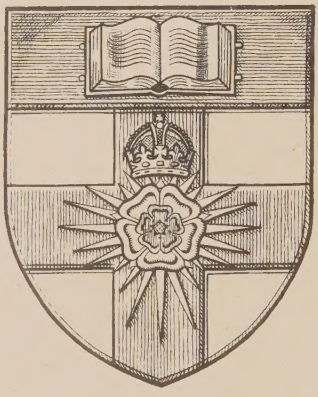
P.S. [illegible] Depository

SHL
WITHDRAWN



UNIVERSITY OF LONDON
LIBRARY

Digitized by the Internet Archive
in 2026

https://archive.org/details/british-journal-of-psychology_1952_5

PR G2

THE BRITISH JOURNAL OF PSYCHOLOGY

Statistical Section

EDITED BY
CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ARTHUR SUMMERFIELD
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
B. BABINGTON SMITH
W. STEPHENSON
P. E. VERNON
J. W. WHITFIELD

Managing Editor : H. G. MAULE

Volume V
1952

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4.

BIBL.
LONDIN.
UNIV.

Vol. V
Part I

The British Journal of Psychology,
Statistical Section

March,
1952

THE FUNDAMENTAL EQUATION OF FACTOR ANALYSIS

By H. KESTELMAN

University College, London

This article is the outcome of a discussion between the author, Sir Cyril Burt, and Mr. A. Summerfield on the basic principles of factor analysis. It is primarily concerned with the mathematical solution of an algebraic problem. But a few preliminary words are needed to show how the problem itself lies at the foundation of nearly all factorial procedures.[1]

Practically every writer on factor analysis assumes that, in the initial stages at least, the object of the method is to express n sets of correlated measurements (which, for brevity, we may call ' tests ') as linear functions of q sets of uncorrelated measurements (the so-called ' factors '), both measurements being in unitary standard measure. Thus, given an $n \times N$ matrix of test-measurements S, we desire to determine an $n \times q$ matrix of weights M, and a $q \times N$ matrix of factor-measurements P, such that $S = MP$ and $PP' = I$.[2]

The solution originally suggested by Karl Pearson ((1) p. 569, (2) p. 209), and later advocated by Hotelling and others (3), was to apply the familiar principal axis transformation. In that case M and P may be found by solving the ' characteristic equation ' of SS', and then determining its ' characteristic ' or ' latent ' roots, V, and the corresponding ' latent vectors,' L. We can then take $P = V^{-\frac{1}{2}}LS = WS$ (say). Since W is a square non-singular matrix, it will have an inverse ; and we thus obtain $S = W^{-1}P$, so that $M = W^{-1}$ ((4) pp. 78–80). The solution thus obtained is both exact and unique.

This procedure yields the same number of factors as tests ($q = n$) ; and all the factors are ' common factors.' But in applying factorial procedures to psychological data, most psychologists have considered it justifiable to seek only those common factors that have some claim to statistical significance and at the same time to allow for the influence of ' specific factors,' each peculiar to a single test. Under these conditions it seems generally assumed that no exact solution can be found : we can, it is said, only calculate what are called (perhaps a little loosely) ' estimated values ' for the factor-measurements ; and these ' estimated values ' will not, after all, prove to be uncorrelated, as the initial statement of the problem required.

Thomson, for example, writes : " The Hotelling components . . . *can be calculated exactly* from a man's scores, . . . whereas Spearman or Thurstone factors can only be estimated. This is because the Hotelling components are never more numerous than the tests, whereas the Thurstone or Spearman factors, including the specifics, are always more numerous than the tests. For the Hotelling components, therefore, we always have just the same number of equations as unknowns, whereas we have more unknowns than equations in the Spearman-Thurstone system " ((8) p. 78). And again " there is no need to ' estimate ' Hotelling's factors ; they can be calculated without any loss of exactness because they are equal in number to the tests . . . Spearman or Thurstone factors, however, are more numerous than the tests, and can therefore only be ' estimated.' Two men with the same set of test scores may have different Thurstone factors " (p. 102). Holzinger

[1] The following statement of the problem is taken from a note by Professor Burt. He adds that it is important that relations between mathematical *variables* should not be confused with relations between statistical *variates*, and points out that the study of the former may be an important preliminary to the study of the latter. It is with the former that this paper is primarily concerned.

[2] This double formulation of factorial requirements appears (in a different notation) in nearly all text-books on the subject ; e.g., (3) p. 428 ; (5) p. 258 ; (7) p. 71, eq. (2), and p. 72, eq. (11) ; (8) p. 358, eq. (3) and (5).

and Harman similarly state : " Since the total number of factors, both common and unique, exceeds the number of variables, the value of any particular factor for a given individual cannot be obtained by direct solution, but can only be estimated from the observed values of the variables " ((6) p. 265). As regards the relations between the factor-measurements, Thomson writes : " We said that one advantage (of using factors) is that the factors are independent and uncorrelated. So they are if their true values are known. *But we only know their estimates, and these estimates are correlated*" ((8) p. 118, his italics). Other writers, echoing these statements, even give the impression that some unattainable set of measurements, at once exact and unique, actually ' exists '—the ' true values,' and that the data only enable one to make more or less satisfactory approximations to these pre-existent ' true values.' Indeed, the distinction between unique and exact values is often overlooked ; and it is therefore desirable to emphasize at the outset that, even when the uncorrelated values are exact (i.e., furnish determinate solutions to the given equations), they need not necessarily be unique.

It will be the purpose of this paper to prove that, even when we assume that the factors are more numerous than the tests, exact numerical specifications can still be found for each of the factors, such that the factor-measurements obtained will be in standard measure and entirely uncorrelated. Of course, it does not follow that such measurements will necessarily be superior from a statistical standpoint. There may be theoretical or practical grounds for preferring the correlated estimates calculated by the equations in current use (e.g., the fact that they have been so determined as to give the ' best ' fit, judged by the principle of least squares). But these are further issues with which this article is not concerned.

To a mathematician it would seem essential to begin by examining, and if possible solving, the prior question here raised, particularly as hitherto it does not seem to have been considered in any direct or explicit way.[1]

1. Let us suppose then that N persons have been subjected to a battery of n tests and that the score of the jth person in the ith test is the (i, j)th element of a matrix,[2] ${}_nS_N$. Then, if the measurements for each test are standardized, i.e., they have zero mean and variance N^{-1}, one of the end products of factor analysis is an equation of the type

$$ {}_nS_N = {}_nM_q \, P_N, \tag{1.1} $$

where q is some number greater than or equal to n. The matrices M and P are restricted to satisfy certain statistical requirements ; if we regard P as the matrix of scores of the N persons in q ' hypothetical tests,' we may require, for example, that the hypothetical test-measurements shall be standardized and uncorrelated. If we denote by I_q the $q \times q$ matrix with unity in the principal diagonal and zero elsewhere, and by ${}_N\sigma_1$ the $N \times 1$ column vector all of whose elements are unity, these requirements are expressed by the equations

$$ {}_qP_N \, P_q' = I_q, \tag{1.2} $$

and

$$ {}_qP_N\sigma_1 = 0. \tag{1.3} $$

Plainly, if (1.1) and (1.2) are satisfied, then

$$ {}_nM_q \, M_n' = {}_nS_N \, S_n'. \tag{1.4} $$

The main object of this paper is to prove a converse, namely, that for *any* M which satisfies (1.4) there are matrices P which verify (1.1), (1.2), and (1.3). The results are here stated and proved formally as exercises in matrix algebra ; their statistical implications are only touched on lightly.

It may, of course, be objected that the statistical requirement of uncorrelated factor-measurements ought not to be given first priority. Professor M. S. Bartlett

[1] Since this paper was written, Professor Kendall has drawn attention to much the same problem. The psychologist, he points out, introduces " a specific term " peculiar to each observed variable, and " this involves him in the very serious trouble that his equations are then essentially indeterminate, a difficulty which has not been satisfactorily resolved " ((9) p. 18).

[2] ${}_pA_q$ denotes a matrix with p rows and q columns.

has pointed out that a situation may arise in which different persons with identical test scores *cannot* be given identical factor scores if (1.2) is to be satisfied. This will not happen if M is square and non-singular, for then $P = M^{-1}S$; and consequently equal columns in the P matrix will correspond with equal columns in the S matrix. But, when $q > n$, the matrix P will not be uniquely determined by S and M, and the situation referred to by Professor Bartlett may then very possibly occur. Suppose, for example, that the N individuals fall into n classes, and that all those in the same class have identical test scores : then, if we require equal test scores to give rise to equal factor scores, the matrix ${}_qP_N$ cannot have more than n different columns, and hence its rank cannot exceed n : but this, when $q > n$, becomes inconsistent with the demand that the factor-measurements for the sample shall be uncorrelated.

2. The data composing S being empirical, there is no loss of generality in assuming that $n < N$ and that S has rank n. If (1.1) holds, then M cannot have rank less than n, and since M has n rows its rank must be exactly n. Hence $q \geqslant n$. If $q = n$, as in the methods used by Pearson, Hotelling, and others, the solution presents no difficulty. Any *square* matrix M which satisfies (1.4) is non-singular, and consequently determines a *unique* matrix P satisfying (1.1), viz., $P = M^{-1}S$; this verifies (1.2), since

$$(M^{-1}S)(M^{-1}S)' = M^{-1}SS'M^{-1'} = M^{-1}(MM')M^{-1'} = I,$$

and it verifies (1.3) because $S\sigma = 0$. The problem of solving (1.1) subject to (1.2) and (1.3) is thus reduced to that of solving (1.4) ; and it is known that the solution of (1.4), when M is square, is

$$M = M_0\Omega, \tag{2.1}$$

where M_0 is a particular solution, and Ω is any orthogonal matrix.

When $q > n$, as in the methods preferred by writers who introduce 'specific' factors, then any ${}_nM_q$ which satisfies (1.4) will also satisfy (1.1) for infinitely many P which do *not* satisfy (1.2). This is because the equation ${}_nM_qx_1 = 0$ has non-trivial solutions, and consequently there are matrices X such that

$${}_nM_q\, X_N = 0 \text{ and } X \neq 0. \tag{2.2}$$

Hence, if $P = \Pi$ satisfies (1.1), so does $P = \Pi + \lambda X$ for any real number λ ; but $P = \Pi + \lambda X$ satisfies (1.2) only if

$$\Pi\Pi' + \lambda(X\Pi' + \Pi X') + \lambda^2 XX' = I ;$$

thus, if $P = \Pi$ satisfies (1.2), then $P = \Pi + \lambda X$ does not (except possibly for one particular non-zero λ).

Now if M satisfies (1.4) it has rank n ; and so there *are* corresponding matrices P verifying (1.1) : there are in fact infinitely many such P ; and our task is to prove that some of them satisfy (1.2) and (1.3).

It is a natural conjecture [1] that the solution of (1.4) for given S and q must have an analogue to the solution (2.1) which applies when $n = q$, i.e., when M is square. This is indeed so; and it opens the way to a solution of (1.1) subject to the conditions (1.2) and (1.3).

3. *Notation.* If the first q columns of ${}_rM_s$ form a matrix ${}_rA_q$ and the remaining $(s - q)$ columns form ${}_rB_{s-q}$, we write $M = (A, B)$. ${}_r0_s$ is the zero matrix of r rows and s columns. If a collection of vectors are mutually orthogonal and each has unit length, we say they are an orthonormal set.

Proposition I. If $n \leqslant N$, a necessary and sufficient condition for a real matrix

[1] Cf. Burt, C. 'Correlation between Persons,' *Brit. J. Psych., Gen. Sect.*, XXVIII, 1937, pp. 84f.

${}_nX_N$ to satisfy the equation $XX' = I_n$ is that X consists of the first n rows of an orthogonal matrix ${}_N\Omega_N$, i.e., that $X = {}_n(I_n, {}_n0_{N-n}){}_N\Omega_N$.

Proof. Sufficiency is trivial. Now suppose ${}_nX_NX_n' = I_n$; this means that the rows of X, say, $\omega_1, \omega_2, \ldots \omega_n$ are an orthonormal set of N-dimensional row vectors. They can therefore be supplemented (if $n < N$) by $\omega_{n+1}, \ldots, \omega_N$ so that $\omega_1, \ldots, \omega_N$ is an orthonormal set. The matrix Ω whose rows are $\omega_1, \ldots, \omega_N$ is then orthogonal and its first n rows compose X.

Correlation. If $n \leqslant N$ and ${}_nX_NX_n' = {}_nY_NY_n' = I_n$ (X and Y being real), then

$$Y = X\Omega_N \text{ for some orthogonal } \Omega. \tag{3.1}$$

This follows at once from Proposition I, for if $X = (I_n, 0)U_N$ and $Y = (I_n, 0)V_N$ where U and V are orthogonal, then $\Omega = U^{-1}V$ is orthogonal and verifies (3.1).

Proposition II. Suppose ${}_nS_N$ is a real matrix of rank n. If q is any integer satisfying $n \leqslant q \leqslant N$, then a necessary and sufficient condition for a real matrix ${}_nX_q$ to satisfy

$${}_nX_qX_n' = {}_nS_NS_n' \tag{3.2}$$

is that $$({}_nX_q, {}_n0_{N-q}) = S\Omega_N \text{ for some orthogonal } \Omega\ ; \tag{3.3}$$

(in other words $${}_nX_q = {}_nS_NV_q, \tag{3.4}$$

where ${}_NV_q$ consists of the first q columns of an orthogonal matrix Ω the remaining $N - q$ columns of which are vectors x satisfying $Sx = 0$).

Proof. If (3.3) holds, then $XX' = (S\Omega)(S\Omega)' = SS'$, i.e., (3.2) holds. Next, suppose that (3.2) holds : we have to verify (3.3). Since SS' is the matrix of a positive-definite quadratic form in n real variables, there is a real diagonal matrix D_n and an orthogonal matrix U_n such that

$$UXX'U' = USS'U' = D^2, \text{ and } |D| \neq 0. \tag{3.5}$$

If we define ${}_qT_N$ to be $(I_q, {}_q0_{N-q})$, i.e., the first q rows of I_N, then $TT' = I_q$, and (3.5) may be written as

$${}_n(D^{-1}UXT)_N\,(T'X'U'D^{-1})_n = {}_n(D^{-1}US)_N\,(S'U'D^{-1})_n = I_n. \tag{3.6}$$

By the corollary to Proposition I, there is an orthogonal matrix ${}_N\Omega_N$ satisfying

$$D^{-1}UXT = D^{-1}US\Omega.$$

Hence $$XT = S\Omega,$$

which means the same as (3.3).

Proposition III. Suppose ${}_nS_N$ has rank n and ${}_nM_q$ is *any* real matrix satisfying

$${}_nM_qM_n' = SS' \text{ and } n \leqslant q \leqslant N\ ; \tag{3.7}$$

then matrices ${}_qP_N$ can be found such that

$$S = MP \text{ and } PP' = I. \tag{3.8}$$

Proof. By Proposition II,

$${}_nM_q = {}_nS_NV_q, \tag{3.9}$$

where ${}_NV_q$ consists of the first q columns of an orthogonal matrix Ω, say $({}_NV_q, {}_NW_{N-q})$, such that $SW = 0$. By (3.9),

$$MV' = SVV' = S(VV' + WW') = S\Omega\Omega' = S.$$

Since the rows of V' are columns of an orthogonal matrix Ω, it follows that $P = V'$ satisfies (3.8).

The solution of (3.8), given S and M, is not unique, for if we have one solution ${}_qP_N{}^*$, and Ω_N is any orthogonal matrix, then $P^*\Omega$ is also a solution of (3.8) if $S\Omega = S$,

that is, if each row x of S satisfies $x\Omega = x$: in other words every row of S is a latent vector of Ω' belonging to the latent root 1.

Proposition IV. Suppose ${}_nS_N$ is a real matrix of rank n, and the sum of the elements in each row of S is zero. Suppose also that $m \geqslant 0$ and

$$S = {}_nM_{m+n}\Phi_N \text{ with } \Phi\Phi' = I_{m+n}. \tag{3.10}$$

Then, *without altering* M, we can find matrices ${}_{m+n}\Phi_N$ such that

$$S = {}_nM_{m+n}\,\Phi_N, \quad \Phi\Phi' = I,$$

and the sum of the elements in any row of Φ is zero.

Proof. The row space of S has dimension n and, by (3.10), it is part of the row space of Φ, which has dimension $m + n$. It follows that we may choose an orthonormal set of N-dimensional row vectors, $\omega_1, \omega_2, \ldots, \omega_N$ such that $\omega_1, \omega_2, \ldots, \omega_n$ span the row space of S and $\omega_1, \omega_2, \ldots, \omega_{m+n}$ that of Φ.

If σ is the unit row vector $N^{-\frac{1}{2}}(1, 1, \ldots, 1)$, it follows from the hypothesis on the rows of S that σ is orthogonal to each of $\omega_1, \omega_2, \ldots, \omega_n$. Let $\xi_1, \xi_2, \ldots, \xi_N$ be any orthonormal set of N-dimensional row vectors such that $\xi_r = \omega_r$ ($r = 1, 2, \ldots, n$) and $\xi_N = \sigma$. Define ${}_NU_N$ by the equation

$$\begin{bmatrix} \omega_1 \\ \omega_2 \\ \cdot \\ \cdot \\ \cdot \\ \omega_N \end{bmatrix} U = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_N \end{bmatrix};$$

then U is orthogonal since the other matrices are. Since $\omega_r U = \omega_r$ ($r = 1, 2, \ldots, n$), and $\omega_1, \omega_2, \ldots, \omega_n$ span the row space of S, it follows $vU = v$ for every row vector v of S. Hence

$$S = SU = M\Phi U.$$

Now let $\Phi = \Phi U$; then $S = M\Phi$ and $\Phi\Phi' = I$. It remains to verify that $\Phi\sigma' = 0$. Since $\omega_r U = \xi_r$ for $1 \leqslant r \leqslant m + n$, and each of these ξ_r is orthogonal to ξ_N, which is σ, it follows that $vU\sigma' = 0$ for every v linearly dependent on $\omega_1, \ldots, \omega_{m+n}$. Hence $\Phi U\sigma' = 0$, i.e., $\Phi\sigma' = 0$.

4. Propositions III and IV imply the result stated in Section 1, viz.,

Theorem. If ${}_nS_N$ is a standardized score matrix and ${}_nM_{m+n}$ is *any* matrix such that MM' is equal to the correlation matrix SS', then, if $n \leqslant m + n \leqslant N$, we can find *standardized and uncorrelated* factor-measurements which define a matrix P such that $S = MP$.

The factor-measurement matrix P will not be unique for a given S and M ; but the correlations between test- and factor-measurements *are* determined by M because $SP' = MPP' = M$. The matrix M which is restricted by the condition $MM' = SS'$ may be further restricted by assigning a pattern : for example, when Pearson's pattern was modified to include factors peculiar to each test, it became

$${}_nM_{m+n} = \begin{bmatrix} f_{11} & f_{12} & \ldots & f_{1m} & s_1 & & \\ \cdot & \cdot & \ldots & \cdot & & s_2 & \\ \cdot & \cdot & \ldots & \cdot & & & \cdot \\ f_{n1} & f_{n2} & \ldots & f_{nm} & & & s_n \end{bmatrix} = ({}_nF_m, {}_nD_n),$$

which denotes the existence of m common factors and n specific ones. The condition $MM' = SS'$ then takes the form

$${}_nF_mF_n' = SS' - D^2. \tag{4.1}$$

For a given S, the diagonal matrix D can be arbitrarily chosen subject to $SS' - D^2$ remaining positive semi-definite. If the choice of D makes $SS' - D^2$ of rank m, then

$$SS' - D^2 = {}_n\Omega_n \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_m & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \Omega',$$

where Ω is orthogonal and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the positive latent roots of $SS' - D^2$. If $\omega_1, \ldots, \omega_m$ are the first m columns of Ω, this means

$$SS' - D^2 = {}_n(\lambda_1^{\frac{1}{2}}\omega_1, \lambda_2^{\frac{1}{2}}\omega_2, \ldots, \lambda_m^{\frac{1}{2}}\omega_m)_m \begin{bmatrix} \lambda_1^{\frac{1}{2}}\omega_1' \\ \cdot \\ \cdot \\ \cdot \\ \lambda_m^{\frac{1}{2}}\omega_m' \end{bmatrix}_n ;$$

hence the m mutually orthogonal vectors $\lambda_1^{\frac{1}{2}}\omega_1, \ldots, \lambda_m^{\frac{1}{2}}\omega_m$ are the columns of a matrix ${}_nF_m$ which satisfies (4.1). Consequently by the Theorem there will be a matrix P of *standardized and uncorrelated* factor-measurements such that $S = (F, D)P$.

REFERENCES

1. Pearson, K. (1901). ' On lines and planes of closest fit to systems of points in space.' *Phil. Mag.*, 6 ser., II, 559–72.
2. Macdonell, W. R. (1902). ' On criminal anthropometry and the identification of criminals.' *Biometrika*, I, 177–227.
3. Hotelling, H. (1933). ' Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, XXIV, 417–41, 498–520.
4. Burt, C. (1937). ' Correlations between persons,' *Brit. J. Psych., Gen. Sect.*, XXVIII, 59–96.
5. Burt, C. (1940). *Factors of the Mind.* London : University of London Press.
6. Holzinger, K. J., and Harman, H. H. (1941). *Factor Analysis.* Chicago : University of Chicago Press.
7. Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago : University of Chicago Press.
8. Thomson, G. (1948). *Factorial Analysis of Human Ability.* London : University of London Press.
9. Kendall, M. G. (1951). ' Regression, structure, and functional relationship.' *Biometrika*, XXXVIII, 11–25.

Vol. V
Part I

The British Journal of Psychology,
Statistical Section

March,
1952

FACTOR ROTATION BY THE METHOD OF EXTENDED VECTORS

A REVIEW OF DR. SUTHERLAND'S PAPER

By T. RENSHAW
Trinity College, Dublin

I. *Introduction.* II. *Criticism of Sutherland's Analysis.* III. *Revised Solution.*
IV. *Summary and Conclusions.*

I. INTRODUCTION

In a recent article (1), Dr. J. Sutherland has described an analysis into oblique factors, using Thurstone's method of extended vectors, but allowing psychological judgement rather than the principle of ' simple structure ' to determine the positions of the reference axes. For his particular battery he claims to have located, in a single series of rotation, the best possible positions of the reference axes. That the resulting factors are psychologically meaningful is, he assumes, a sufficient criterion for the acceptance of the oblique solution. It should be noted that Thurstone himself does not rule out the use of psychological judgement in making the rotations, for he writes :

' In exploratory factorial studies it is often found that the primary factors are not completely determined in some of the diagrams because the test batteries in exploratory work seldom reveal a complete simple structure. It is then advisable to use all the information available, including the configuration as well as the psychological nature of the tests, to determine the co-ordinate planes, which are then interpreted as in the nature of hypotheses concerning the underlying parameters or primary factors ' ((2) p. 256).

It is the view of the present writer that, while Dr. Sutherland has shown much skill in locating the reference axes, he has not given adequate attention to the nature of the resulting structure. By not carefully inspecting the diagrams which show the configuration of test points relative to his chosen reference axes, he has failed to obtain the most plausible explanation of the factors involved. In this paper, it is proposed to point out what adjustments in his oblique solution are required by the configuration, and to suggest a revised analysis of his data. The revised solution will, it is thought, prove of particular interest in that it indicates a possible reconciliation of several theories concerning the structure of mental abilities.

II. CRITICISM OF SUTHERLAND'S ANALYSIS

An unsatisfactory feature of Sutherland's oblique solution is the ' schooling ' factor, which correlates negatively to the extent of ·870 with his oblique number factor and negatively to the extent of ·898 with his oblique verbal factor ((1) p. 27, Table VI). If we plot, in pairs, the columns of his reference vector structure ((1) p. 26, Table V), several linear arrangements of points become evident, which suggest further adjustments to the positions of the reference vectors. The most striking of these are illustrated in Figs. 1 (*a*) and 1 (*b*) (where unit distance from the origin is indicated by an arrow-head).

Each of these diagrams gives evidence of desirable adjustment to factor IV_1 (schooling). Let us consider their outcome. The adjustment suggested in Fig. 1 (*a*) would rotate factor IV_1 into coincidence with factor III_1 (verbal), while that suggested

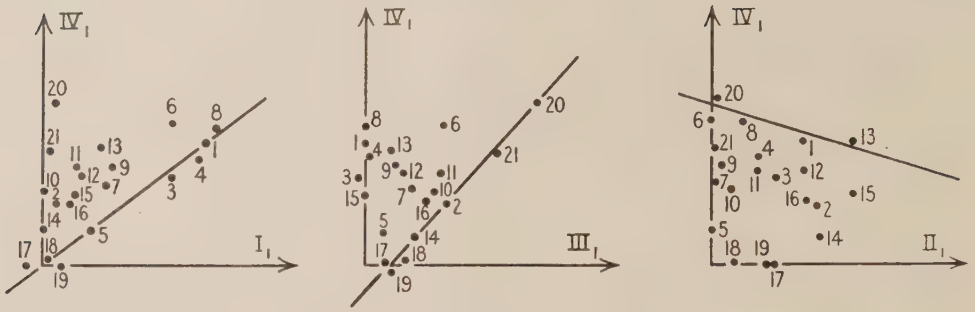


FIG. 1.—To illustrate methods of rotating centroid factors.

in Fig. 1 (*b*) would rotate factor IV_1 into coincidence with factor I_1 (number); the result in each case is the loss of one dimension of the configuration. What we have illustrated geometrically is the mathematical fact that vectors I_1, III_1, and IV_1 are *linearly dependent*.

Now it is a necessary condition that the reference vectors should be *linearly independent* ((2) p. 238); and since vectors I_1 and III_1 are both well defined, it becomes necessary to re-determine vector IV_1. An alternative position is clearly indicated in Fig. 1 (*c*); and it is interesting to note that virtually the same line is apparent in Figs. 1 (*a*) and 1 (*b*).

III. REVISED SOLUTION

The Primary Factors. The rotation suggested in Fig. 1 (*c*) was accordingly carried out, with further adjustments indicated by the configuration. Table I shows the resulting simple structure.[1]

TABLE I.· SIMPLE STRUCTURE ON THE REFERENCE VECTORS: $F\Lambda$

No.	Test	A	B	C	D	E
1	Arithmetic Quotient	—·069	·520	·331	·001	·171
2	I.Q. (Picture Test)	·076	—·014	·345	—·052	·426
3	Teachers' Arithmetic Estimate	—·063	·513	·215	·219	·206
4	Mechanical Arithmetic	—·038	·617	·166	·048	·199
5	Problem Arithmetic	·056	·414	—·059	·471	·287
6	Mechanical Arithmetic	·219	·523	—·016	—·058	·138
7	Problem Arithmetic	·166	·352	—·014	·341	·182
8	Speed Mechanical Arithmetic	·004	·634	·129	·247	—·004
9	Speed Problem Arithmetic	·196	·306	·005	·390	·009
10	Spatial Problem Arithmetic	·270	·074	·015	·406	·117
11	Verbal Intelligence	·225	·151	·126	·265	·207
12	Non-Verbal Intelligence	·148	·015	·313	·307	·082
13	Jenkins Test (A)	·108	—·033	·498	·263	—·009
14	Matrix Test	—·012	—·009	·333	·063	·433
15	Jenkins Test (B)	·003	—·035	·472	·373	·105
16	Space Test (A)	·051	·026	·322	—·058	·319
17	Space Test (B)	·001	·025	·160	·360	·327
18	Number Series	·019	·249	·012	·362	·470
19	Letter Series	—·062	·232	·136	·334	·483
20	Vocabulary	·551	—·048	·016	—·040	·037
21	Interpretation	·401	·037	—·005	·055	·155

[1] For reasons of space it is not possible to give details of the calculations. They are available in the editorial office for anyone who may wish to see them.

Factor A (verbal), B (number), and C (space) are interpretations that agree with Sutherland's. Factor E is reminiscent of Thurstone's ' Inductive ' factor ((3) p. 86), later renamed ' Reasoning ' (4). Having no first-hand acquaintance with the test battery, the present writer cannot make more than a superficial attempt at interpreting factor D. The results appear to indicate a reasoning factor in which the problem arithmetic tests have high loadings. Such a factor has been reported by previous investigators.[1]

The cosines of the angles between the primary factors were calculated by the formula : $D(\Lambda'\Lambda)^{-1}D$. They are given in Table II. It will be noticed that all the primary factors are positively correlated.

TABLE II. CORRELATIONS BETWEEN THE PRIMARY FACTORS : $D(\Lambda'\Lambda)^{-1}D$

Factor	A	B	C	D	E
A	1·000	·583	·625	·336	·494
B	·583	1·000	·320	·085	·069
C	·625	·320	1·000	·428	·303
D	·336	·085	·428	1·000	·269
E	·494	·069	·303	·269	1·000

The Second-Order Solution. Two second-order factors were found sufficient to account for the inter-correlations between the primary factors, close agreement between inserted and resulting communalities being obtained after six iterations. Contrary to Sutherland's analysis, and for a reason that will appear later, the second-order factors were not rotated at this stage. Instead the unrotated loadings were entered in matrix E which was completed as shown in Table III, the sum of the squares in each row being unity. The last column gives the entries for the diagonal matrix D^{-1}.

TABLE III. SECOND-ORDER ANALYSIS : MATRIX E

	I	II	S_1	S_2	S_3	S_4	S_5	Multipliers : D^{-1}
A	·971	—·208	·120					1·7329
B	·496	—·492		·716				1·3038
C	·701	·118			·704			1·3449
D	·525	·497				·691		1·1269
E	·424	·085					·902	1·2161

The matrix product $F\Lambda D^{-1}E$ was next formed. This yields an analysis of the original twenty-one tests into two orthogonal second-order factors *plus* an orthogonal simple structure (Table IV).

Agreement between obtained and centroid communalities is, on the whole, quite good. As Sutherland points out, perfect agreement can be expected only when the correlations between the primary factors are analysed completely. A case similar to that recorded by Sutherland ((1) p. 28, footnote 2), in which a communality greater than unity was obtained, has been observed by the present writer. It resulted during

[1] E.g., by Burt, by Davis, Guilford, and Lacey (quoted by Vernon (5) pp. 55–6), and by Thurstone ((3) p. 88), and more recently by Barakat (this *Journal*, IV, pp. 65f.).

TABLE IV. SECOND-ORDER FACTORS *PLUS* ORTHOGONAL SIMPLE STRUCTURE: $G = F\Lambda D^{-1}E$

No.	I	II	I_1	II_1	1	2	3	4	5	Communalities	
										Obtained	Centroid
1	·620	—·237	·468	·471	—·015	·485	·313	·001	·188	·809	·798
2	·633	·051	·598	·214	·016	—·013	·327	—·041	·467	·731	·741
3	·663	—·128	·552	·389	—·013	·479	·203	·170	·226	·806	·804
4	·622	—·308	·440	·536	—·008	·576	·157	·037	·218	·887	·862
5	·733	—·001	·668	·302	·012	·387	—·056	·367	·314	·923	·923
6	·728	—·434	·486	·695	·046	·488	—·015	—·045	·151	·984	·970
7	·788	—·078	·687	·395	·035	·329	—·013	·266	·200	·847	·847
8	·682	—·249	·519	·507	·001	·592	·122	·193	—·004	·929	·890
9	·767	—·047	·680	·357	·041	·286	·004	·304	·010	·766	·767
10	·818	·097	·785	·247	·056	·070	·014	·317	·129	·803	·788
11	·858	·012	·788	·342	·047	·141	·119	·206	·227	·868	·865
12	·777	·167	·777	·166	·031	·014	·296	·239	·090	·785	·774
13	·781	·207	·797	·131	·023	—·031	·471	·205	—·010	·918	·899
14	·549	·143	·559	·095	—·002	—·008	·315	·049	·475	·649	·646
15	·702	·316	·770	·000	·001	—·033	·446	·291	·115	·891	·886
16	·536	·016	·495	·205	·011	·025	·304	—·045	·349	·505	·507
17	·549	·244	·601	·003	·000	·024	·152	·280	·358	·592	·592
18	·660	·087	·638	·192	·004	·232	·012	·282	·516	·843	·826
19	·621	·132	·620	·135	—·013	·217	·129	·260	·530	·815	·794
20	·907	—·184	·751	·540	·115	—·045	·015	—·031	·041	·874	·887
21	·807	—·123	·685	·443	·084	·035	—·005	·043	·170	·705	·722

an analysis in which minor adjustments indicated by the configuration in the oblique rotations had not been carried out and was corrected after these adjustments were made. This suggests that, whenever we wish to proceed with a second-order analysis, all adjustments to the oblique solution should first be adequately effected.

Interpreting the Second-Order Factors. Second-order factor I (Table IV) has been located by averaging the primary factors ; it is near the centre of the configuration, and might reasonably be interpreted as '*g*.' Second-order factor II, on the other hand, is bipolar and presents a dichotomous grouping similar to Vernon's *v* : *ed* — *k* : *m* classification. This is of particular interest, for it will be observed that no such classification is indicated in the matrix of centroid loadings ((1) p. 21, Table II). There, the pattern of signs for the second centroid axis does not arrange the verbal and numerical tests together. The reader may therefore ask : Does this not demonstrate the arbitrary nature of a classification based on a centroid or bipolar analysis, and does it not lead us to question Burt's method of partitioning the correlation matrix for purposes of group factor analysis (6), and perhaps even to doubt the hierarchical picture of the structure of human abilities favoured by both Burt and Vernon ? Against this, it might be claimed that, although the second centroid factor loadings of the verbal tests in this particular battery differ in sign from the numerical tests, the former loadings are not significantly different from zero. Nevertheless, it can be shown that the difference between the smallest negative loading — ·173 (tests 5 and 7, problem arithmetic) and the smallest positive loading ·069 (test 21, interpretation) borders on significance ($P = ·05$) ; and if we substitute for test 21 the vocabulary test, the difference is fully significant. Again, in Thurstone's Primary Mental Abilities battery, according to the centroid analysis ((3) pp. 113–14), the arithmetic tests do not appear in the *v* : *ed* sub-division as would be expected according to Vernon's classification. In view of these criticisms it is therefore of particular interest to find that the second-order analysis has restored the *v* : *ed* — *k* : *m* dichotomy absent in the centroid analysis.

Second-order factors I and II were next rotated, keeping them orthogonal, in order to make all loadings in factor II positive. The matrix of the transformation, calculated in the usual manner, was :—

$$\begin{bmatrix} ·9119 & ·4104 \\ ·4104 & —·9119 \end{bmatrix}$$

Fig. 2 shows the new factorial axes I_1 and II_1, and the loadings in these factors are given in Table IV.

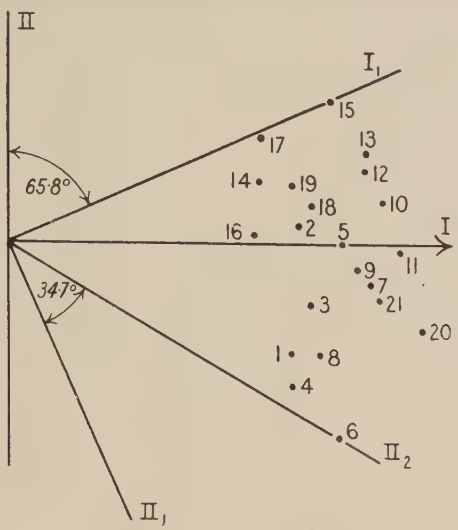


FIG. 2.—To illustrate rotation of second-order factors.

Axis I_1 has been placed through test points 15 and 17, and factor II_1 is interpreted as 'schooling,' the assumption being that performance on tests 15 (Jenkins B) and 17 (Space B) is not influenced by schooling. In that case factor I_1 would be a general factor independent of schooling. But is this the best solution suggested by the configuration of test points? Is it not desirable to place axis II through test point 6? When this is done, we have two oblique second-order factors, I_1 and II_2; and these, in the opinion of the writer, may well be interpreted as '*g-k*' and '*g-ed.*' The transformation matrix for this further adjustment, which moves only one of the second-order factors, is :—

$$\begin{bmatrix} 1 & \cdot 5693 \\ 0 & \cdot 8221 \end{bmatrix}$$

In the columns of this matrix are entered the direction cosines of the new oblique axes, and the inner product of the columns, namely ·5693, represents the correlation between factors '*gk*' and '*g ed.*'

The correlations between the tests and the oblique second-order factors are shown in Table V. This table also gives the correlations between the tests and the primary factors. The formula for this latter calculation is given by Thomson ((7) p. 375). It is $F(\Lambda')^{-1}D$. For convenience, in our particular case, this formula was transformed as follows :—

$$\begin{aligned} F(\Lambda')^{-1}D &= F\Lambda\Lambda^{-1}(\Lambda')^{-1}D \\ &= F\Lambda D^{-1}D\Lambda^{-1}(\Lambda')^{-1}D \\ &= F\Lambda D^{-1}D(\Lambda'\Lambda)^{-1}D \end{aligned}$$

The matrix product $F\Lambda$ is shown in Table I; the product $D(\Lambda'\Lambda)^{-1}D$ in Table II; and the entries for the diagonal matrix D^{-1} in Table III.

The figures in Table V, being the correlations between the tests and the oblique factors, indicate the validity of each test as a measure of the different factors. It will be seen that test 20 (vocabulary) yields a highly valid estimate of factor A, which we may call verbal ability. All the tests in the battery have a predictive value for this ability; and this result seems plausible. The validity of the mechanical arithmetic tests for factor B is quite high, whereas that of the non-verbal intelligence tests for the same factor is very low. Of this particular group of tests, the one with highest predictive value for the space factor is Jenkins Test (A) and the same test provides the best measure of '*gk.*' Estimates for factors D and E obtainable with the tests of this battery are of lower validity than those for the other factors; but it is conceivable that tests of higher validity could be constructed. As a measure of '*g ed,*' the vocabulary test has highest validity; and it will be observed that the

TABLE V. STRUCTURE ON THE SECOND-ORDER AND PRIMARY FACTORS

No.	Test	I_1	II_2	A	B	C	D	E
1	Arithmetic Quotient	·468	·653	·656	·764	·650	·265	·331
2	I.Q. (Picture Test)	·598	·516	·648	·238	·673	·322	·707
3	Teachers' Arithmetic Estimate	·552	·634	·667	·735	·616	·457	·396
4	Mechanical Arithmetic	·440	·691	·680	·858	·535	·261	·347
5	Problem Arithmetic	·668	·628	·713	·640	·487	·669	·553
6	Mechanical Arithmetic	·486	·848	·824	·902	·457	·157	·378
7	Problem Arithmetic	·687	·715	·781	·668	·539	·571	·493
8	Speed Mechanical Arithmetic	·519	·713	·688	·909	·560	·424	·183
9	Speed Problem Arithmetic	·680	·681	·729	·637	·537	·592	·326
10	Spatial Problem Arithmetic	·785	·650	·762	·425	·584	·671	·510
11	Verbal Intelligence	·788	·729	·836	·521	·681	·586	·590
12	Non-Verbal Intelligence	·777	·579	·696	·339	·765	·641	·448
13	Jenkins Test (A)	·797	·562	·675	·305	·896	·640	·361
14	Matrix Test	·559	·397	·537	·162	·621	·396	·670
15	Jenkins Test (B)	·770	·438	·580	·205	·842	·725	·432
16	Space Test (A)	·495	·451	·548	·245	·588	·255	·546
17	Space Test (B)	·601	·344	·487	·164	·520	·607	·575
18	Number Series	·638	·521	·651	·422	·488	·607	·725
19	Letter Series	·620	·464	·601	·372	·552	·602	·712
20	Vocabulary	·751	·872	·939	·500	·592	·373	·507
21	Interpretation	·685	·754	·833	·470	·526	·347	·550

same test has a validity for factor '*gk*' approaching that of the non-verbal intelligence tests. This finding emphasizes Terman's view that a vocabulary test provides the best single measure of intelligence ((8) p. 302).

That two second-order factors only are required to account for the inter-correlations between the primary factors, and that these can be interpreted as '*g*' and 'schooling,' provides, in the opinion of the writer, convincing evidence in support of Thurstone's theory of correlated factors of ability. The findings, too, are in substantial agreement with Thomson's sampling theory, as may be judged by the following quotation from his discussion of the nature of mental structure. "Some of this 'structure' is no doubt innate; but more is probably due to environment and education and life. Its expression in terms of separate uncorrelated factors suggests what is almost certainly not the case, that the 'sub-pools' are separate from one another" ((7) p. 319). It remains to be seen whether further enquiries will confirm the second-order solution outlined above. That two other analyses[1] by the writer have done so suggests that they might.

Finally, the fact that fewer second-order factors were isolated here than in Sutherland's analysis does not conflict with the views concerning the complexity of factors expressed by Thomson in his presidential address to the Psychology Section of the British Association (10). Indeed, as already pointed out, our conclusions agree closely with his sampling theory. Moreover, the problem of 'specifics' raised in the above-mentioned address is here partly overcome in that the communalities in Sutherland's battery are on the average very high.

IV. SUMMARY AND CONCLUSIONS

1. The oblique factorial solution presented by Dr. Sutherland is criticized on the ground that rotations indicated by the configuration of test points were not adequately completed.

[1] One, of a battery of verbal and non-verbal intelligence tests, space tests, and attainment tests in English and Arithmetic administered to eleven-year-old children (9); and the other, a re-analysis of Vernon's battery of similar tests administered to army recruits ((5) p. 23, Table 23).

2. While agreeing with Sutherland that, by carefully locating the reference axes in the initial rotation, it is possible to arrive at a close approximation to the best possible oblique structure, it is contended that the investigator must consider with equal care whether further rotations are suggested in the diagrams showing the positions of the test points relative to the chosen reference axes.

3. Minor adjustments to the selected axes, i.e., small angular rotations of a few degrees, need not be undertaken except when the investigator wishes to proceed to a second-order analysis. In that case it is necessary to ensure that the estimates of the inter-correlations between the primary factors shall be as accurate as possible ; otherwise the rank of the matrix of inter-correlations between the primary factors may be appreciably altered, and an anomalous second-order solution may result.

4. In the light of the above criticisms, a re-analysis of Sutherland's data was undertaken. The revised solution outlined in this paper yields five oblique primary factors, four of which correspond to the oblique verbal, number, spatial, and reasoning factors isolated by Sutherland. The fifth resembles Thurstone's ' Inductive ' factor.

5. Two second-order factors, which may be interpreted as ' *g* ' and ' schooling,' appear sufficient to account for the inter-correlations between the primary factors.

6. Alternative rotations of these second-order factors, oblique and orthogonal, indicate a possible reconciliation between the hierarchical and the correlated factor theories of the structure of human abilities.

7. That fewer second-order factors were isolated than in Sutherland's analysis does not, in the opinion of the writer, conflict with Thomson's views concerning the complexity of the factors so obtained.[1]

REFERENCES

1. Sutherland, J. (1951). ' Factor rotation by the method of extended vectors.' *Brit. J. Psych., Stat. Sect.*, IV, 21–30.
2. Thurstone, L. L. (1947). *Multiple-Factor Analysis.* Chicago : The University of Chicago Press.
3. Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago : The University of Chicago Press.
4. Thurstone, L. L. and T. G. (1941). *Factorial Studies of Intelligence.* Chicago : The University of Chicago Press.
5. Vernon, P. E. (1950). *The Structure of Human Abilities.* London : Methuen.
6. Burt, Sir Cyril (1950). ' Group factor analysis.' *Brit. J. Psych., Stat. Sect.*, III, 40–75.
7. Thomson, Sir Godfrey H. (1948). *The Factorial Analysis of Human Ability.* (3rd ed.) London : University of London Press.
8. Terman, L. M., and Merrill, M. A. (1937). *Measuring Intelligence.* London : Harrap.
9. Renshaw, T. (1950). ' A factorial study of two and three dimensional space tests.' Ph.D. Thesis, University of Edinburgh.
10. Thomson, Sir Godfrey H. (1949). ' The nature of the mind's " factors." ' *The Advancement of Science,* VI, 23, 267–74.

Comment by Dr. J. Sutherland. The editor has kindly allowed me to comment on Dr. Renshaw's interesting article. While I am inclined to agree that Renshaw's solution is probably better than the original, I do not think that it alters the fundamental conclusions.

Throughout his article one has the impression that the writer believes that there must always be a ' best ' solution—a best position for the reference vectors, for example : this may not be the only solution, but it stands out as better than others ; moreover, it will not be simply a matter of opinion or of psychological interpretation : it will make itself evident in the configuration of the test points, and the fact that it is the ' best ' solution will be confirmed by the way in which all the later calculations drop into place. Now, despite the examples quoted by Renshaw, I still doubt

[1] The author is indebted to Sir Cyril Burt and Mr. Arthur Summerfield for helpful comments on his manuscript.

all this. Thurstone himself, despite his hopes for Simple Structure, seems to doubt it. He says : " the purpose of this chapter is to show that, *when a clear and complete simple structure exists* as in the present fictional example, the rotational methods that have been described enable us to identify the underlying order " (*Multiple Factor Analysis*, p. 256, italics mine). I myself have not had a case of simple structure arising in an actual experimental situation ; and I feel that, in our present state of knowledge, psychological interpretation must still be the main guide.

Again, although by means of a second rotation of one pair of reference vectors, Renshaw obtains a better position for Reference Vector IV, it is because he alters the conditions it is required to satisfy —rightly so, if he wants it to be more independent of the others.

The significant difference between Renshaw's correlations for the Primary Factors and mine is, I think, due rather to the fact that the correlations are smaller than that they are all positive. His five primary factors are therefore more nearly orthogonal than mine. This question of independence of factors raises an interesting point in which there seems some inconsistency in the theory. If, in order to obtain a more easily interpreted solution, we decide to use oblique factors, then the reference vectors cannot be completely independent ; and if we wish to place our reference vectors in positions which give the most reasonable psychological factors, then we cannot rule out the possibility of highly correlated factors. On the other hand, if we wish to explain the correlations in terms of the minimum number of factors, we may well prefer to make the factors as independent of each other as possible. If we carry this to its limit, we come back to orthogonal factors.

J. SUTHERLAND.

Note on Bipolar and Group Factor Methods. In his interesting re-analysis of Dr. Sutherland's correlation table, Mr. Renshaw has discovered a broad group factor entering into all the 'Intellectual' tests (including both the Arithmetical and Verbal) and into none of the 'Practical' tests ; and this factor, as he points out, finds no representation in Dr. Sutherland's matrix of centroid factors. " Does not this demonstrate," he asks, " the arbitrary nature of a classification based on a centroid or bipolar analysis ? And does it not lead us to question Burt's method of partitioning a correlation matrix for purposes of group factor analysis, and perhaps even to doubt the hierarchical picture of the structure of human abilities favoured by both Burt and Vernon ? "

There are three questions here, each of which deserves a brief reply. May I take the second question first, that which deals with the " method of partitioning a correlation matrix for purposes of a group factor analysis " ? In the publications to which Mr. Renshaw refers I argued that what I called an ' analysis by submatrices ' would reveal appropriate group factors with a computing technique that was at once quicker and more objective than the graphical method of rotating factors two or three at a time which Thurstone preferred to substitute for it. The procedure involved partitioning the correlation matrix into separate submatrices, corresponding to the ' groups ' of tests ; and for this purpose the starting point adopted was the classification suggested by the sign pattern of the bipolar factors furnished by the preliminary summational analysis. Ideally this summational analysis should be, not a ' centroid analysis,' proceeding by *simple* summation, but one based on *weighted* summation. As I have always insisted, the simplified method which I put forward in 1917 (substituting what has since been called the ' centroid formula ' for Pearson's determinantal formula) is to be regarded only as an approximate device. Nevertheless, save in exceptional cases, the pattern of signs furnished by simple summation differs but little in its relevant features from that furnished by the more laborious procedure.

It is, however, essential to bear in mind what conditions are presupposed if the simplified procedure is to yield valid results. First, the summational analysis must be based on a correlation matrix in which the appropriate reduced self-correlations or 'communalities' have been inserted. This usually entails two or three iterations.[1] Now Thurstone's modification of my summational procedure instructs the computer simply to take for each communality the highest correlation (or residual correlation) in the corresponding column. This simplification provides a quick and mechanical rule which always appeals to the inexperienced computer. But it entails certain marked disadvantages. Weighted summation maximizes the first few factors and minimizes the later. The substitution of simple summation tends to enlarge the later factors and so make them appear more significant than they really are. This tendency is still further increased by inserting the highest correlations in place of the more appropriate figures. The effects may be readily seen in Dr. Sutherland's matrix of centroid loadings (Table II, p. 21). There the fourth factor actually has a higher contribution to the variance than the third. Moreover, we are told that, even after five complete common factors had been extracted, the residuals were still significant. In a table based on only 136 cases, so many significant factors could hardly occur with weighted summation based on reduced self-correlations reached by successive approximation.

[1] Dr. Sutherland explicitly says that his matrix was " factorized by the centroid method, using the highest correlation in each column as communality and without iteration " (*loc. cit.*, p. 21). Since his object was to study the value of Thurstone's procedure by applying it to a specific case, he was bound to follow that procedure implicitly. My criticisms are not intended to reflect on Dr. Sutherland's paper, which provides a valuable example of the detailed methods proposed.

Secondly, if we are content to use simple summation in order to obtain an appropriate classification of the tests, then it is important that the tests should have been selected so as not to hamper this result. With simple summation the saturations necessarily add up to zero : that means that the number of tests in the positive and negative sections or subsections must be fairly equally balanced. Now, out of 21 items in Dr. Sutherland's battery, 10 are Numerical tests and 4 (at most) are Verbal : that makes 14 Intellectual tests and only 7 Practical tests. With these unequal numbers we can hardly expect a centroid analysis to give negative signs to all the 14 Intellectual tests and positive to only 7 Practical tests. The general effect of such a lack of balance is obvious : nearly always the tests from the larger group that have the lowest saturations get their signs reversed in order to maintain an unweighted total of zero. With the present series, after more appropriate self-correlations have been inserted, two of the smallest saturations actually change back from positive to negative ; and, on proceeding to carry out an analysis by weighted summation, we find that the first bipolar factor classifies all the so-called Verbal tests with the Arithmetical tests, and the distinction between the two subgroups only emerges with the next bipolar factor.[1]

This seems to provide a reply to the first as well as to the second of Mr. Renshaw's questions. If my contentions are correct, what is ' arbitrary ' is not ' the classification indicated by the bipolar analysis,' but the number of tests employed to represent the various classes or types.

Finally, Mr. Renshaw believes that a study of Dr. Sutherland's data leads to doubts about " the hierarchical picture of the structure of human abilities favoured by both Burt and Vernon." To answer this suggestion we must examine the group factor matrix obtained by the principle of partitioning that I have put forward. For the results to be convincing, the group factors should be derived quite automatically from the numerical data, without any reference to the actual abilities or the tests with which the data have been obtained. Since I myself was already familiar with the names of the 21 tests and other items, I handed a copy of Dr. Sutherland's tables of correlations and of centroid factors to a former student, Mr. R. F. Morgan, who has had some experience of factorial methods, but gave him no indication as to the nature of the variables. He has been good enough to carry out a group factor analysis on perfectly mechanical lines, and to accompany this result by a note of the various steps in his calculations.

In the table of correlations as actually printed by Dr. Sutherland (*loc. cit.*, Table I, p. 22) the lines of division between the various classes are somewhat obscured by the irregular order in which the items are arranged. The first step in carrying out a group factor analysis is to rearrange the table in accordance with the main divisions indicated by the sign patterns.[2] This means that the Picture test will be placed with the other visual tests instead of between two Arithmetical entries ; the Vocabulary test placed after the other Verbal tests instead of after the Spatial ; and the tests of Mechanical Arithmetic grouped together, instead of alternating with the tests of Problem Arithmetic. When this is done, the general correctness of the method of partitioning I have proposed becomes obvious at once.

The order thus reached is shown in Table VI below, and is based on the arrangement indicated by the bipolar factors resulting from *weighted* summation.[3] These suggest a classification into (A) Intellectual tests (1 to 29 in the rearranged order) and (B) Practical or Spatial tests (2 to 14) ; the former being subdivided into (A1) Numerical tests (1 to 9) and (A2) Relational tests (18 to 29), with smaller subgroups for the purely Verbal tests and one or two other types (Mechanical Arithmetic, Space, etc.).

On inspecting the rearranged correlation table, it becomes evident that most of the overlapping will occur with group A2. The submatrix giving cross-correlations between (A1) the Numerical and (B) the Spatial tests is evidently very nearly hierarchical. These figures can therefore be accounted for by a single factor only, namely, the ' basic factor,' which (when present) is determined first.

So far as the tests in these two groups are concerned, therefore, we can at once determine the saturations for the first or basic factor by the procedure outlined for a group factor analysis with

[1] Mr. Renshaw admits that the positive centroid loadings given by Dr. Sutherland for the Verbal tests are non-significant ; and I myself should argue that this prevents them from being classified with the Practical tests. However, in anticipation of this argument, Mr. Renshaw also claims that " the difference between the smallest *positive* loading and the smallest *negative* loading borders on significance, and that, if we substitute the Vocabulary test for test 21, the difference is fully significant." This, however, does not seem very convincing, especially as he nowhere states what criterion he has employed to determine statistical significance. The standard error for the majority of the correlations in Dr. Sutherland's table is well below 0·04. Hence in what follows a residual has been treated as significant only if it exceeds 0·08 (or an appropriately smaller value for the largest correlations).

[2] *The Factors of the Mind*, p. 477 ; ' Group Factor Analysis,' this *Journal*, III, p. 45.

[3] Even if we provisionally accept the classification indicated by Dr. Sutherland's centroid matrix as a convenient starting-point, the residuals left by a preliminary group factor analysis based on his figures will at once indicate the need for slight modifications. However, with a table like the present, we can, if we prefer, avoid the labour of successive approximation at the preliminary bipolar stage by leaving it to the group factor stage.

two group factors only.[1] Further, except for slight divergences in tests 10 and 19, the observed cross-correlations between (*a*) all the tests in group A2 and (*b*) tests 12, 13, and 16 in Group B also fit into the same hierarchy ; and these figures enable us to calculate the basic saturations for the remaining tests.

After the basic factor has been removed, the residuals in group A can again be factorized by the method appropriate to the case of two group factors, since the residual cross-correlations between groups A1 and A2 are again nearly hierarchical. We now have one basic factor common to all the tests, one broad group factor common to the first 14 tests, and three non-overlapping group factors corresponding to sections A1, A2, and B. The residuals obtained after these have been extracted provide the requisite data for the saturations due to overlap and to the small group factors.[2]

The final results are shown in Table VI. With only two exceptions out of 56, the residuals are all numerically less than $\pm$ 0·080.[3]

Dr. Sutherland was anxious to obtain factors " which could be interpreted psychologically." Accordingly, if the group factor procedure, which proceeds without reference to the nature of the tests, is to claim that its final results are equally satisfactory, then we must first show that the factors which it yields are susceptible of an equally plausible interpretation. Let us therefore briefly glance at each of the factors in turn to see what kind of abilities they indicate.

1. The first factor is plainly a ' basic factor,' which may be roughly identified with ' *General Ability* ' or ' intelligence.' It is not quite the same as the ' general factor ' finally reached by Dr. Sutherland, since, as he points out, that is still " weighted on the arithmetical side." In other words, his method of rotation has not succeeded in eliminating the results of choosing a battery consisting largely of Arithmetical tests. With our own general or ' basic ' factor the highest saturations are those for Verbal Intelligence, Non-verbal Intelligence, Jenkins A, and Problem Arithmetic (in that order). Hence, unlike the general factor obtained by Dr. Sutherland, the basic factor obtained by the group factor method appears to be free from undue bias.

2. The second factor appears identifiable with what in previous publications I have called an *Intellectual* (as distinguished from a Practical) factor. As a rule, it enters, though in varying degrees, into both verbal or literary and arithmetical or mathematical attainments. It depends quite as much on interest as on ability ; in the younger child it plays an important part in predicting what has been called ' educable capacity,' and in the older child reflects the results of the education he has already received and the way in which that has influenced his intellectual interests and attainments. To some extent it resembles what Professor Vernon terms *v* : *ed.* But whether, with any particular group of pupils, it is chiefly verbal or chiefly arithmetical turns largely on the type of schooling they have received. At the elementary or primary stage, where arithmetic is apt to be the crucial subject, it often proves (as it does here) to be weighted more strongly for arithmetical than for verbal subjects. Hence Mr. Renshaw's suggestion that in the present instance it should be termed a Verbal-Educational factor (*v* : *ed*) does not seem to me to be quite convincing. Apparently it has little relation to the factor which Dr. Sutherland calls ' schooling ' (p. 26),[4] since that factor is strongly weighted with the Pictorial and Spatial tests.

3. The next factor is a factor for *Arithmetical* ability and attainments. It enters into all the tests of mechanical and problem arithmetic, as well as the ' arithmetical quotient ' and the ' teacher's assessments for arithmetic.' There is a small but scarcely significant overlap with the test described as ' Number Series ' ; and a still smaller non-significant overlap with ' Spatial Problem Arithmetic ' (not entered in the table, since Mr. Morgan excluded instances where the saturation would fall below 0·100 : if required, these numerous small instances of non-significant overlap could be dealt with much better by an arithmetical rotation).

4. The fourth factor bears a close resemblance to the factor called ' verbal ' in Dr. Sutherland's Table VIII and ' reasoning ' in his Table III. In my 1917 paper I found two verbal factors : an

[1] This *Journal*, *loc. cit. sup.*, pp. 53f.

[2] Where there is much overlapping, a method of successive approximation is sometimes necessary. (Here only the earlier saturations for the few instances of overlap and of doublets and triplets required adjustment.) Where the amount of overlapping is doubtful, it is wiser to proceed to an arithmetical rotation, based on the preliminary group factor analysis : (cf. *loc. cit.*, pp. 61f., and appendix on ' Calculation of Rotation Matrix,' p. 75 : an excellent instance of this further procedure will be found in Mr. Barakat's recent article on " A Factorial Study of Mathematical Abilities," this *Journal*, IV, pp. 145f.).

[3] The two exceptions are for tests 6 and 15, — ·241 ; tests 5 and 17, — ·123 : (the first is more than three times the standard error of the observed correlation, but the correlation itself is somewhat anomalous—only ·273). Since both residuals are negative, they cannot be attributed to ordinary overlapping. However, one such case might easily arise by chance in a table of 56 coefficients.

[4] I am not, however, altogether clear why Dr. Sutherland's reference vector IV_1 is called ' schooling.' It has no negative saturations at all, and enters into practically every one of his tests. Consequently I should have thought it was really identifiable with the general factor.

elementary ' word factor ' entering largely into Vocabulary tests for reading, spelling, defining, where the examinee has to deal merely with isolated words ; and a ' language ' (or ' literary ') factor which might be described *Verbal thinking* ; it figures largely in literary tests of English composition and tests of non-literary subjects where the answers are of an essay type. It plays a considerable part in the solution of problems which are presented in numerical or spatial form but are solved by the aid of verbal imagery. I myself should have identified the fourth factor that emerges from my re-analysis of Dr. Sutherland's data with this factor for ' verbal thinking ' ; but it may equally well be a relational factor. And Dr. Sutherland, it seems, would prefer to describe it as a factor for ' reasoning.' Accordingly, I have provisionally labelled it ' *R*,' since he alone has first-hand knowledge of his tests. I would only remark that ordinarily it is extremely difficult to discriminate a special ability for reasoning (or using logical relations) from general intelligence itself unless the battery is appropriately constructed : it was perhaps unfortunate that the present battery included no tests expressly designed to distinguish the higher verbal or literary factor from the factor of reasoning, or that introspections were not available to check the interpretations.

5. The last of the larger group factors seems plainly to be a *Spatial factor*. It is almost identical with the factor termed ' space ' in three of Dr. Sutherland's tables.

6. The more elementary *Word factor* evidently plays an important part in the last two tests in Dr. Sutherland's list, namely, the Vocabulary test and what he calls ' Interpretation.' It would seem to be identical with the verbal reference vector mentioned in his Table V, which has large correlations for this pair of tests only. Curiously enough it appears to overlap with test 6 (Mechanical Arithmetic 2) ; and the overlapping also appears in Dr. Sutherland's tables. With my own analysis, however, the figure is barely significant statistically, and is difficult to interpret without a detailed knowledge of the problems included in test 6.

7. The four remaining factors arise from the repetition of the same or similar tests ; but do not appear in the analyses of either Dr. Sutherland or Mr. Renshaw. Nevertheless, the residuals on which they are based are fully significant ; and a glance at the initial correlation table (when appropriately rearranged) at once suggests their presence. There is first of all a narrow group factor entering into the three tests of *Mechanical Arithmetic*. This shows some slight overlapping with the ' Teachers' Assessment for Arithmetic ' and with the ' Arithmetical Quotient,' though here the saturations are scarcely significant. It is a factor which has appeared quite frequently in previous investigations : there it seems largely based on mechanical memory and a mechanical knowledge of arithmetical tables and of the four rules resulting from memorization.

8. The two *Jenkins tests* provide the highest coefficient in the table of observed correlations, except that for the two tests of Mechanical Arithmetic (·900 and ·904 respectively). These correlations seem to me to be really of the nature of reliability coefficients.[1] However, if the tests are inserted as separate items in the initial table, then these correlations must inevitably introduce a small group factor.

9. There is also a significant residual left from the correlation between the two ' *Space tests*,' which again seems rather of the nature of a reliability coefficient.

10. Finally there is a significant residual for the two Speed tests (8 and 9). The existence of a *Speed factor* has been found in a number of previous researches.

The factors thus obtained therefore all seem to lend themselves quite readily to an intelligible psychological interpretation ; and, on the whole, in spite of Mr. Renshaw's doubts, the factor matrix fits in reasonably well with the hypothesis of hierarchical structure outlined in my previous review.[2]

Accordingly, I venture to maintain that my method of obtaining group factors is preferable to that proposed by Professor Thurstone and followed by Dr. Sutherland and Mr. Renshaw. In the first place, it is much easier and speedier to carry out. Secondly, it is far more objective : if the rotations themselves are guided by the investigator's views about what interpretations are psychologically preferable, then the results cannot be claimed as a proof of those views ; for, as Mr. Renshaw quite rightly remarks, there are numerous alternative positions which can be interpreted psychologically, and different investigators are free to select those which agree best with their own preconceptions. Thus, if (with Mr. Renshaw) he believes in the existence of the *v* : *ed* factor, he is likely to find it ; if he does not, it will not be difficult to avoid it. If he believes in a general or basic factor, it will appear ; if not, the use of correlated factors or of freely overlapping factor-saturations can easily conceal it. The fact that Dr. Sutherland himself reaches so many different interpretations seems to me to indicate the somewhat precarious nature of the method he has adopted. Thus his Table III giving the results of his first rotation reveals a general factor

[1] I might perhaps point out that the group factor analysis would have been very much simpler if, in accordance with the usual procedure, repetitions of the same tests had been pooled.

[2] ' The Structure of the Mind : A Review of the Results of Factor Analysis,' *Brit. J. Educ. Psychol.*, XIX, 1949, pp. 100–99 ; cf. *The Measurement of Mental Capacities* (Oliver and Boyd, 1935), pp. 31–48. The five broader factors mentioned above are all to be found in this early publication and the remainder (except the factor for the two Jenkins tests) are found in the later.

and four group factors. Table V apparently includes no general factor, but introduces a new factor of schooling. Table VIII displays eight factors, including two reasoning factors, a schooling factor, and a factor of interest.[1]

Thirdly, the method I have adopted reveals three or four narrow group factors, which are not discovered by the methods adopted by Dr. Sutherland and Mr. Renshaw, but whose existence can scarcely be doubted. These are the factors entering chiefly into the doublets, the two Jenkins tests, the two Space tests, the two Speed tests, and the two or three tests of Mechanical Arithmetic.

Mr. Renshaw's final results (two general factors, with all group factors reduced to a single factor of schooling), so far from providing 'convincing evidence for Thurstone's theory,' seem to me, when taken in conjunction with Dr. Sutherland's, merely to show how inconclusive the method adopted can be.

CYRIL BURT

TABLE VI. SATURATIONS OBTAINED BY GROUP FACTOR ANALYSIS

Test	Factor									
	G	I	N	R	K	V	Mech. Ar.	Spe.	J.	Spa.
1. Arithmetic Quotient ..	·736	·212	·325				(·288)			
3. Arithmetic Teachers' Est.	·779	·320	·288				(·099)			
4. Mechanical Arithmetic (A)	·666	·456	·348				·301			
6. Mechanical Arithmetic (B)	·697	·428	·356			(·138)	·312			
8. Mechanical Arithmetic (Speed)	·561	·472	·537				·423	·313		
5. Problem Arithmetic (A)	·807	·407	·138	(·181)						
7. Problem Arithmetic (B)	·782	·422	·094	(·212)						
9. Problem Arithmetic (Speed)	·739	·398	·136	(·148)				·313		
18. Number Series ..	·715	·269		·468						
19. Letter Series	·659	·317		·489						
10. Spatial Problem Arithmetic..	·698	·324		·359						
11. Verbal Intelligence ..	·844	·246		·360						
20. Vocabulary	·778	·178		·141		·654				
29. Interpretation	·727	·207		·206		·316				
2. Picture Test	·719			(·189)	·354					
12. Non-verbal Intelligence	·841				·237					
13. Jenkins Test (A) ..	·821				·387				·420	
15. Jenkins Test (B) ..	·738			(·168)	·323				·420	
16. Space Test (A) ..	·576				·330					·434
17. Space Test (B).. ..	·632			·325	·249					·434
14. Matrix Test	·619			·241	·362					

[1] As he points out, the factor which is termed 'interest' produces (with two minor exceptions where the saturation is virtually zero) negative signs for all the Arithmetical tests and positive signs for all the Non-arithmetical tests. One would have supposed that the natural interpretation of this would be to regard it as a bipolar factor distinguishing arithmetical from other abilities. The reasons for interpreting it as a factor of 'interest or novelty' are not very clear.

In the rotated matrices published by those who adopt "the guiding principle" that, at each stage, axes are to be rotated to positions which the investigator can interpret psychologically, there is one common feature which several critics have noted. During the earlier stages the investigator tends to fix first on positions that he himself prefers, and then ends up with several supplementary factors (often more numerous than the centroid matrix would appear to justify) whose nature tends to be at once miscellaneous and obscure, and which include a large proportion of variance, since they have been forced to take up such portions of it as would have been incorporated in factors that he does not willingly favour.

Vol. V
Part I

The British Journal of Psychology,
Statistical Section

March,
1952

A FACTORIAL ANALYSIS OF ITEMS IN THE BERNREUTER PERSONALITY INVENTORY

By CHARLOTTE BANKS and GERTRUDE KEIR

Department of Psychology, University College, London

I. *Problem.* (*a*) *Origin and Construction of Scale.* (*b*) *Results of Previous Investigations.* (*c*) *Need for a Factor Analysis of Items.* II. *The Present Investigation.* (*a*) *Aim.* (*b*) *Subjects.* (*c*) *Method.* (*d*) *Results.* III. *Summary.*

I. PROBLEM

(*a*) *Origin and Construction of Scale.* The Bernreuter Personality Inventory (2) was originally designed to measure four traits—neurotic tendency, self-sufficiency, introversion, and dominance. The complete version consists of 125 items in question form, each of which is answered by encircling 'Yes,' 'No,' or '?.' Bernreuter considered that his inventory "represents a new departure in that it measures different aspects of personality at the same time." This is rendered possible by using four schemes of differential weighting, or 'scales,' one for each of the four traits. As a result, says Flanagan, "this inventory is one of an increasing number which are unique [1] in that they permit *several* scores to be obtained from *one* set of items" ((11) p. 38). Bernreuter describes the traits assessed by his Inventory somewhat more fully as follows:

(i) 'B1–N' is "a measure of *Neurotic tendency*; persons scoring high in it tend towards emotional instability."

(ii) 'B2–S' is "a measure of *Self-sufficiency*; persons scoring high in it prefer to be alone and rarely seek sympathy or encouragement."

(iii) 'B3–I' is "a measure of *Introversion-extraversion*; persons scoring high in it are imaginative and tend to live within themselves."

(iv) 'B4–D' is "a measure of *Dominance-submission*; persons scoring high in it tend to dominate others in face to face situations."

Flanagan has factorized the inter-correlations between the total scores for each of the four traits, and found two independent factors only, namely, Self-confidence and Sociability ((11) p. 48). These were subsequently incorporated by Bernreuter in his revised manual, and termed F1–C and F2–S. F1–C is a measure of *lack of confidence*: persons scoring high on this scale "tend to be hamperingly self-conscious and to have feelings of inferiority." F2–S is a measure of *lack of sociability*: persons scoring high "tend to be non-social, solitary, or independent" (4).

[1] Although, at the time that Bernreuter and Flanagan wrote, the notion of obtaining scores for a *multiplicity* of factors from a *single* 'battery' was comparatively new in American research, it had been in use in this country almost from the time when Karl Pearson introduced the 'method of principal axes' for deducing a number of 'index characters' from one and the same set of measurements. The device had chiefly been used in the educational field, e.g., for distinguishing pupils suited for a grammar school from those suited for technical schools, and to discriminate special forms of educational abilities; it was first applied to the field of 'personality' (in the narrower sense of affective and conative traits) in Burt's paper on 'General and special factors underlying the primary emotions' (*Brit. Assoc. Ann. Rep.*, 1915, pp. 694f.). Bernreuter's method of weighting was not based on any preliminary factorial study: Flanagan, however, attempted to derive a practical working procedure, based on an analysis into principal components ((11) pp. 49f.).

The questions of which Bernreuter's Inventory is composed were taken in the main from four previously constructed scales—Thurstone's Neurotic Inventory,[1] Laird's C2 Introversion test, Allport's test of Ascendance-Submission, and Bernreuter's own test of Self-sufficiency. These were first administered to batches of American college students, each being given to about 350 to 450 men and women. From these larger batches four criterion groups were selected, consisting of the 25 men and 25 women who had obtained the highest and the lowest scores for each trait. On the basis of the responses given by these selected groups to Bernreuter's composite questionnaire, ' diagnostic values ' were then assigned to the several items. This was done by the aid of the Strong chart and Kelley's formula for determining the weights $\left(\frac{\varphi}{(1-\varphi^2)\ \sigma}\right)$. Final weights, ranging from -7 to $+7$, are assigned to each item for each of the four traits. Bernreuter prints a selected list of the ten items most highly weighted for each of the four, namely, B1–N, B2–S, B3–I, and B4–D respectively (2).

(*b*) *Results of Previous Investigations.* Reliability coefficients for the four main traits have been given both by Bernreuter and by later investigators. As a rule the coefficients have been calculated by the split-half method, corrected by the Spearman-Brown formula.[2] The figures generally range from 0·85 to 0·90.

To determine the validity of the four traits, Bernreuter applied the inventory to two small groups of psychology students, numbering 70 and 55 respectively. The criteria consisted of scores obtained in the four original tests (those named above) from which the inventory had been constructed. In a section on ' Validity ' ((4) p. 4) he gives correlations, with and without correction for attenuation, based on these results. The uncorrected coefficients range from ·67 to ·91. But such a procedure hardly constitutes an adequate validation.[3]

Other researches on the validity of the scales present a less optimistic picture. Thus Landis and his co-workers found that none of the Bernreuter scales would differentiate groups of normal students from groups of psychoneurotic and psychotic patients (13, 14); and Mosier believes that " their conclusion concerning the validity of the test is probably the best guide to its use." He adds that this does not necessarily detract from the value of the information that may be derived from an examinee's replies to questions separately considered ((5) p. 51).

[1] In Thurstone's questionnaire the heading actually used is ' Personality Schedule.' This does not mean that the schedule was intended to cover all aspects of personality : the phrase was used to avoid the reluctance which some examinees might feel over answering questions designed to expose neurotic weaknesses.

[2] The Spearman-Brown formula for a split-half reliability seems particularly unsuited to such scales as the present : the resulting coefficients differ widely according to the way in which the lists are split. Burt has suggested a more general ' variance formula,' which is virtually the average of the reliabilities obtained by all possible splits, namely,

$$r_{tt'} = \frac{n}{n-1}\left\{1 - \frac{\Sigma\sigma^2}{\sigma_t^2}\right\}$$

(*Brit. J. Educ. Psychol.*, XV, 1945, pp. 91–2, eq. (xiii) and (xiv)). With this formula we find the reliabilities of the several scales are : B1–N, 0·74 ; B2–S, 0·69 ; B3–I, 0·78 ; B4–D, 0·76. These figures are nearer to Flanagan's ((11) p. 42) than to Bernreuter's.

[3] First, many of the items in the Bernreuter inventory were taken from these other tests—a fact bound to increase the correlations between these tests and the Bernreuter. Secondly, the practice of validating one inventory against another is not very sound, since inventories of this sort usually show rather low correlations with ratings based on actual behaviour. Thus, with the groups on which the tests were standardized, Bernreuter's Self-sufficiency Scale and Allport's Ascendance-Submission Scale yield correlations with independent ratings amounting to only about ·50 ; and drop to about 0·30 with other groups. The differences in the figures may perhaps be due partly to the fact that the different groups are variously selected, and hence have different means and standard deviations for the traits in question. Indeed, a good deal of the work by Bernreuter and others on the inventory has been criticized because the subjects have so often consisted of selected groups, such as students. To meet such criticisms Burt has suggested that a correction for selection could be incorporated in the formulæ for correlation : in that way invariant parameters could be obtained, which would be independent of differences due to selection : (*Psychometrika*, IX, pp. 219f.). However, in an exploratory study like the present it seemed scarcely necessary to complicate the calculations with these further refinements.

Bernreuter also gives figures showing the inter-correlations between the six traits scored according to his weighting scales ((4) p. 5). The coefficients are based on replies from 157 engineering students (men only). As he points out, at least two of the traits overlap considerably (the correlations between B1–N and B3–I being ·95—a figure higher than the reliability coefficient for either); and several of the other pairs are inversely related (the correlations of B4–D with B1–N and B3–I being —·80 and —·69 respectively). Similar results have been obtained by subsequent investigators, e.g., Flanagan (11), Lorge (15), and Stagner (18). Commenting on these figures, Newcomb observes that " for all practical purposes it must be concluded that B1–N, B3–I, and B4–D are alternative measures of the same thing, whatever it may be called " ((5) p. 82).

Flanagan has derived two modified weighting scales from Bernreuter's original version, and contends that these might well be substituted for Bernreuter's four. " It would seem," he writes, " that scores on the two new variables would contain such a large amount of the available information that it would not be worth while to score the blanks for more than these two factors " ((11) p. 73). He has calculated the correlations between scores for Bernreuter's four traits as estimated from his own scales (F1–C and F2–S) and the scores obtained by using Bernreuter's scales; the coefficients are as follows: B1–N, ·970; B2–S, ·867; B3–I, ·954; B4–D, ·867. In a later investigation, Sartain (17) calculated regression equations for predicting B1–N and B4–D from F1–C and F2–S, the multiple correlations being as high as ·893 and ·784.

In his review of the literature dealing with the Bernreuter Inventory, Super (19) states that, from the time of its publication down to the end of 1940, more than 135 studies of the test have been reported. During the last 10 years three or four investigations have appeared annually. These figures suggest that Bernreuter's questionnaire must be among the most widely used in this field. For this reason a more thorough enquiry into the nature of the traits which it actually measures seems urgently needed. But such an enquiry should be based on a detailed investigation of the several items, not, like the foregoing studies by Flanagan and others, simply on the total scores.

(*c*) *Need for a Factor Analysis of Items.* Before we can consider how to derive satisfactory assessments for the several traits which the scales claim to measure, it seems clear that we ought first of all to verify the assumption that the questions used to measure ' dominance,' ' introversion,' ' self-sufficiency,' and the like do form *relatively independent* groups, showing (to borrow Karl Pearson's terms) higher correlations between ' like ' than between ' unlike ' questions, or whether they may not be explained in terms of even fewer fundamental traits. This should form an essential stage in the validation of all such scales. For this purpose a factor analysis of the inter-correlations between separate items would seem to be the most appropriate technique. As regards other tests of the questionnaire type, factorial studies of separate items have been attempted by Burt for an early questionnaire of his own (6), by Mosier for Thurstone's questionnaire (16), and by J. P. and R. B. Guilford for ' typical items ' from questionnaires on introversion (12). But nothing of the sort seems hitherto to have been undertaken for Bernreuter's Inventory.

II. THE PRESENT INVESTIGATION

(*a*) *Aim.* Our primary aim was to determine the number and nature of the factors underlying the Bernreuter Inventory. However, the labour involved in analysing a correlation matrix based on 125 variables would make a factorial procedure almost impracticable for the complete set. An obvious simplification is to rely solely on the most typical or most diagnostic items. In the following enquiry, therefore, we shall restrict ourselves to those selected by Bernreuter as most diagnostic. For each trait he gives a list of the 10 most heavily weighted items. But eight of these are common to two of the lists. Hence the total number of questions amounts to only 32. The questions are reprinted in full in the appendix (Table III, p. 30). The headings show the traits for which the items are most heavily weighted in Bernreuter's scales; and the last three columns, the frequency of the responses given by our own subjects.

The course here followed was also adopted by Mosier in his study of Thurstone's Inventory which contained as many as 132 items. Mosier selected 39 of the most diagnostic, and applied Thurstone's method of factor analysis to these. As it happens, 15 of the questions picked out by Mosier are also to be found in Bernreuter's selected lists.

TABLE I. INTER-CORRELATIONS BETWEEN MOS

Item	No.	87*	103*	94*	125*	40*	96*	118*	54	72	5	2
Introduces people at a party	87*		·44	·48	·48	·16	·22	·56	−·37	−·32	−·20	−
No difficulty talking to stranger	103*	·44		·56	·39	·29	·48	·63	−·06	·05	−·18	−
Takes lead at dull party	94*	·48	·56		·50	·52	·40	·61	−·17	−·18	−·22	−
Not self-conscious starting discussion	125*	·48	·39	·50		·61	·21	·38	−·27	−·20	−·24	−
Has organized teams or groups	40*	·16	·29	·52	·61		·14	·15	−·21	−·38	−·10	−
Not reluctant to meet important people	96*	·22	·48	·40	·21	·14		·44	·03	−·45	−·32	−
Not in background on social occasions	118*	·56	·63	·61	·38	·15	·44		−·49	−·24	−·25	−
Feels lonely with others	54	−·37	−·06	−·17	−·27	−·21	·03	−·49		·61	·53	
Troubled by inferiority	72	−·32	·05	−·18	−·20	−·38	−·45	−·24	·61		·46	
Hurt by criticism	5	−·20	−·18	−·22	−·24	−·10	−·32	−·25	·53	·46		
Shy	24	−·33	−·35	−·21	−·56	−·27	−·35	−·56	·54	·57	·39	
Does not like to be with others much	121	−·51	−·10	−·40	−·13	−·07	−·27	−·47	·25	·03	·19	
Could be absorbed in creative work	23	·02	−·30	−·11	−·31	−·01	−·15	−·42	·18	−·05	·13	
Faces troubles alone	110*	·08	−·20	−·11	−·04	−·20	·09	−·19	·13	−·07	·16	
Would take a chance alone	82*	·24	·10	−·04	·10	·42	·10	·21	·28	·36	−·09	−
Understands problems best alone	79*	·10	−·20	−·14	−·15	−·27	−·05	−·08	−·23	·15	·16	
Does not get views before making decision	89*	·24	·17	·03	−·11	·25	·05	−·26	·20	−·14	·09	
Prefers making hurried decisions alone	112*	·11	·33	−·11	·01	−·10	−·10	·07	·04	·21	·26	−
Prefers to work alone	101*	·17	−·25	−·10	·14	·00	−·15	−·27	·24	·17	−·37	
Finds reading more helpful than conversation	38*	−·09	−·16	−·16	−·10	−·16	−·24	−·13	·25	·15	·04	
Enjoys an evening alone	61*	−·22	−·28	−·12	·11	·08	−·06	−·24	·07	·15	−·10	
Moody	71	·17	·06	·04	·15	·23	−·02	·05	·49	·47	·34	
Bothered by some useless thought	73	·03	·19	−·10	·13	−·23	·33	·01	·40	·44	·30	
Worries over possible misfortunes	106	−·20	−·08	−·37	−·25	−·19	−·08	−·28	·49	·45	·46	
Nervous	14	−·31	−·23	−·15	−·31	−·14	−·23	−·33	·45	·65	·54	
Feelings alternate without reason	104	−·03	−·19	·39	−·22	−·08	·01	−·03	·42	·30	·36	
Worries over humiliation	39	·10	−·22	·02	−·30	−·20	−·27	−·10	·43	·47	·34	
Often feels just miserable	8	−·13	−·18	−·32	−·15	−·26	−·03	−·20	·62	·46	·47	
Daydreams frequently	2	·02	−·15	−·18	·10	·13	·12	−·21	·49	·30	·13	
Feelings easily hurt	51	·05	−·13	−·17	−·41	−·20	·01	−·10	·45	·48	·42	
Self-confident	80*	·30	·06	·12	·31	·25	·21	·18	−·42	−·68	−·25	−
Blushes frequently	12	−·20	·12	·01	−·42	−·09	−·07	−·23	·37	·68	·25	

* Items marked by an aster

(*b*) *Subjects.* The entire inventory was administered to 100 students of psychology, comprising 61 men and 39 women, attending a second-year course in personality testing at University College. Their median age was 22·5 years. The inventory was given at the beginning of the course, before any description or discussion about such scales had been started. At this stage nearly all the students were quite unfamiliar with tests of this type.

(*c*) *Method.* The coefficients were computed by the tetrachoric procedure. This implies that, for any given question, the personal quality assumed to determine each person's reply might be regarded as a continuous variable having an approximately normal distribution. Thus it is assumed that ability to 'stand criticism' is a matter of degree ; and that those who can stand criticism with either a moderate or a considerable degree of ease would tend to answer 'Yes,' that those who found it neither decidedly easy nor decidedly hard would probably mark the query, and finally that those who were conscious of a moderate or considerable amount of difficulty in so doing would tend to answer 'No.' To secure a fourfold table, wherever a student marked the query, his reply was grouped either with the affirmative or with the negative replies, whichever arrangement gave a dichotomy nearest to the median.

(*d*) *Results.* No significant differences were found between the coefficients for the two sexes. The correlations obtained for the whole group are shown in Table I. The coefficients range from −0·68 to +0·61. The table was factorized by Burt's Method of Simple Summation with reduced self-correlations. It will be noted that, when Bernreuter's scoring is used, practically all the questions relating to Dominance and to Self-sufficiency show negative correlations with the questions in the other sections. For the sake of consistency, therefore, before the factorial procedure was applied, the direction of the scoring was reversed for those traits

DIAGNOSTIC ITEMS IN THE PERSONALITY INVENTORY

121	23	110*	82*	79*	89*	112*	101*	38*	61*	71	73	106	14	104	39	8	2	51	80*	12
−·51	·02	·08	·24	·10	·24	·11	·17	−·09	−·22	·17	·03	−·20	−·31	−·03	·10	−·13	·02	·05	·30	−·20
−·10	−·30	−·20	·10	−·20	·17	·33	−·25	−·16	−·28	·06	·19	−·08	−·23	−·19	−·22	−·18	−·15	−·13	·06	·12
−·40	−·11	−·11	−·04	−·14	·03	−·11	−·10	−·16	−·12	·04	−·10	−·37	−·15	·39	·02	−·32	−·18	−·17	·12	·01
−·13	−·31	−·04	·10	−·15	−·11	·01	·14	−·10	·11	·15	·13	−·25	−·31	−·22	−·30	−·15	·10	−·41	·31	−·42
−·07	−·01	−·20	·42	−·27	·25	−·10	·00	−·16	·08	·23	−·23	−·19	−·14	−·08	−·20	−·26	·13	−·20	·25	−·09
−·27	−·15	·09	·10	−·05	·05	−·10	−·15	−·24	−·06	−·02	·33	−·08	−·23	·01	−·27	−·03	·12	·01	·21	−·07
−·47	−·42	−·19	·21	−·08	−·26	·07	−·27	−·13	−·24	·05	·01	−·28	−·33	−·03	−·10	−·20	−·21	−·10	·18	−·23
·25	·18	·13	·28	−·23	·20	·04	·24	·25	·07	·49	·40	·49	·45	·42	·43	·62	·49	·45	−·42	·37
·03	−·05	−·07	·36	·15	−·14	·21	·17	·15	·15	·47	·44	·45	·65	·30	·47	·46	·30	·48	−·68	·68
·19	·13	·16	−·09	·16	·09	·26	−·37	·04	−·10	·34	·30	·46	·54	·36	·34	·47	·13	·42	−·25	·25
·30	·36	·08	−·51	·07	·46	−·02	·10	·02	·10	·02	·00	·35	·45	·13	·58	·48	·06	·47	−·62	·64
	−·05	·39	−·10	·18	·20	·08	·38	·28	·24	−·10	−·04	−·10	·10	·01	−·26	−·03	−·10	−·21	−·01	·00
−·05		·03	−·13	·08	·14	·11	−·01	·06	−·13	·07	·19	·17	·26	·50	·36	−·04	−·04	·07	−·08	·15
·39	·03		·16	·56	·36	·55	·45	·35	·39	−·07	·27	−·14	−·29	·21	·09	·21	−·14	−·22	−·03	−·08
−·10	−·13	·16		·19	·23	·20	·23	·42	·12	·00	·07	·10	−·09	·08	−·35	−·20	−·24	−·42	·36	−·69
·18	·08	·56	·19		·26	·40	·17	·44	·49	−·02	−·07	·19	·04	−·31	·08	−·10	−·10	−·12	·05	−·06
·20	·14	·36	·23	·26		·21	·07	·34	·14	−·10	−·14	−·08	−·17	−·27	−·23	·05	−·08	−·23	−·09	−·07
·08	·11	·55	·20	·40	·21		·04	−·16	·17	·06	−·18	−·16	·10	−·18	·07	−·21	−·04	·21	·10	·10
·38	−·01	·45	·23	·17	·07	·04		·23	·08	·16	−·05	·03	·25	·08	−·14	−·14	·09	−·11	·01	·18
·28	·06	·35	·42	·44	·34	−·16	·23		−·06	−·25	−·07	·19	·10	−·31	·03	·00	−·10	·02	−·14	−·30
·24	−·13	·39	·12	·49	·14	·17	·08	−·06		·08	−·05	−·16	−·03	−·18	−·07	−·13	−·10	−·08	·19	−·18
−·10	·07	−·07	·00	−·02	−·10	·06	·16	− 25	·08		·64	·74	·43	·74	·55	·50	·49	·55	−·23	·50
−·04	·19	·27	·07	−·07	−·14	−·18	−·05	−·07	−·05	·64		·57	·52	·53	·44	·62	·52	·50	−·30	·33
−·10	·17	−·14	·10	·19	−·08	−·16	·03	·19	−·16	·74	·57		·45	·33	·45	·66	·26	·45	−·52	·63
·10	·26	−·29	−·09	·04	−·17	·10	·25	·10	−·03	·43	·52	·45		·66	·44	·56	·34	·67	−·68	·64
·01	·50	·21	·08	−·31	−·27	−·18	·08	−·31	−·18	·74	·53	·33	·66		·40	·49	·47	·52	−·35	·49
−·26	·36	·09	−·35	·08	−·23	·07	−·14	·03	−·07	·55	·44	·45	·44	·40		·55	·47	·62	−·44	·32
−·03	−·04	·21	−·20	−·10	·05	−·21	−·14	·00	−·13	·50	·62	·66	·56	·49	·55		·48	·67	−·35	·33
−·10	−·04	−·14	−·24	−·10	−·08	−·04	·09	−·10	−·10	·49	·52	·26	·34	·47	·47	·48		·30	−·30	·60
−·21	·07	−·22	−·42	−·12	−·23	·21	−·11	·02	−·08	·55	·50	·45	·67	·52	·62	·67	·30		−·48	·28
−·01	−·08	−·03	·36	·05	−·09	·10	·01	−·14	·19	−·23	−·30	−·52	−·68	−·35	−·44	−·35	−·30	−·48		−·42
·00	·15	−·08	−·69	−·06	−·07	·10	·18	−·30	−·18	·50	·33	·63	·64	·49	·32	·33	·60	·28	−·42	

re reversed in Table II.

for which the column totals were negative. The factor-saturations are given in Table II ; and the items for which the scores have been reversed are indicated with an asterisk. To obtain the saturations for Bernreuter's scoring the reader has simply to reverse the sign of the coefficients for each of the traits so marked.

Factor I. The first factor is a general factor, with positive signs throughout. It contributes only 23·2 per cent. to the total variance and 54·0 per cent. to the common factor variance—proportions decidedly lower than those usually found with cognitive tests. The highest saturations (over 0·60) are for the following items : 14. *Nervous.* 8. *Often miserable.* 51. *Feelings easily hurt.* 106. *Worries over possible misfortunes.* 39. *Worries over humiliation.* 24. *Shy.* 12. *Blushes.* 80. *Lacks self-confidence.* 104. *Feelings alternate from happy to sad without reason.* 72. *Troubled by feelings of inferiority.* 54. *Feels lonely with others.*

With one exception these questions are all included by Bernreuter either in the Neurotic Scale (B1–N) or in the scale for Introversion (B3–I), and most of them are included in both. It would be a mistake, however, to suppose that the factor therefore represents something that might be termed ' introverted neuroticism.' The general characteristic indicated by these answers would seem to be simply that of ' nervousness ' in the popular sense. A high factor-measurement merely implies that the examinee is of a shy, diffident, sensitive, and somewhat anxious type.

In his analysis of the total scores obtained from the four scales, Flanagan finds a somewhat similar factor, accounting for " about 78 per cent. of the non-chance variance."[1] It has " large positive weights for Neurotic tendency and Introversion (0·89 and 0·86), a rather large weight of the opposite sign for Dominance (−0·83), and a somewhat smaller negative weight for Self-sufficiency

[1] The table analysed was the 4×4 correlation table of ' estimated correlations,' i.e., correlations " obtained in such a way as to eliminate the effect of correlation between errors." The analysis was carried out by the method of principal components with reliability coefficients in the leading diagonal.

(−0·59)." He suggests that " Factor One may be designated by the single word *self-confidence*," by which he seems to mean *lack* of self-confidence ((11) pp. 43, 48).

The general factor found by our analysis of the separate items bears a close resemblance to that obtained by Mosier in his centroid analysis of Thurstone's Neurotic Inventory—a factor which (so we calculate from his table) contributes nearly 25 per cent. to the total variance ((16) p. 282, Table III). Mosier himself offered no interpretation of his centroid factors, but rotated them, by Thurstone's procedure, so as to secure a ' simple structure ' of eight ' primary ' factors. He himself goes on to declare that ' there is no single trait of neurotic tendency which can be postulated in a parsimonious description of behaviour.' The wording is here a little ambiguous. If the statement means that neurotic characteristics cannot be explained *solely* in terms of a single factor, then every psychologist would agree. If, however (as the context rather implies) his real meaning is that there is *no general factor* to be found in the items he analysed, then the results of his own centroid analysis plainly refute

TABLE II. FACTOR-SATURATIONS

Number	Item	Bernreuter's Grouping	Factors I	II	III
87*	Does not introduce people at a party	−D	·415	·454	·288
103*	Has difficulty in talking to strangers	−D	·328	·469	·255
94*	Does not take lead at parties	−D	·365	·568	·220
125*	Self-conscious in starting discussion	−D	·500	·431	·208
40*	Has not organized teams or groups	−D	·356	·291	·181
96*	Reluctant to meet important people	−D	·276	·444	·101
118*	Keeps in background on social occasions	−D	·473	·643	·024
54	Feels lonely with others	N	·617	·191	−·474
72	Troubled by inferiority	N & −D	·618	·118	−·417
5	Hurt by criticism	N & I	·555	·158	−·214
24	Shy	N & −D	·668	·389	−·120
121	Does not like to be with others	+S	·030	·541	−·037
23	Could be absorbed in creative work	+S	·275	·123	·000
110*	Does not face troubles alone	−S	·137	−·423	·552
82*	Not willing to take a chance alone	−S	·399	−·178	·425
79*	Understands problems best in discussion	−S	·151	−·420	·413
89*	Seeks many views before making decision	−S	·163	−·365	·389
112*	Does not prefer to make hurried decisions alone	−S	·102	−·213	·381
101*	Prefers to work with others	−S	·003	−·205	·255
38*	Finds conversation more helpful than reading	−S	·027	−·412	·210
61*	Does not enjoy an evening alone	−S	·113	−·348	·148
71	Moody	I	·494	−·432	−·517
73	Bothered by useless thoughts	N & I	·487	−·359	−·487
106	Worries over possible misfortunes	N	·676	−·086	−·474
14	Nervous	I	·780	−·021	−·285
104	Feelings alternate without reason	N & I	·621	−·314	−·269
39	Worries over humiliation	N & I	·652	−·138	−·217
8	Often feels miserable	N & I	·727	−·152	−·210
2	Daydreams frequently	I	·477	−·330	−·169
51	Feelings easily hurt	N & I	·698	−·244	−·140
80*	Lacks self-confidence	−D	·645	−·006	−·137
12	Blushes frequently	I	·697	−·172	−·131

* Items marked by an asterisk have been reversed in this table.

any such conclusion.[1] The items having the highest saturations with the general factor in his investigation are on the whole similar to those noted above, namely, those relating to inferiority, lack of self-confidence, worry, and misery. But the question relating explicitly to nervousness had a comparatively low saturation.

The fact that the proportion contributed by the general factor to the total variance is slightly higher in Mosier's investigation confirms what might be inferred from an inspection of the content

[1] It may be noted that Burt and two of his research students factorized both Guilford's table of correlations and that of Mosier (reconstructed) by the group factor method, and found a definite general or ' basic ' factor, in addition to group factors (cf. (10) p. 263).

of his questions, namely, that they overlapped with one another still more widely. It is further reflected in the range and size of the saturations found in his table as contrasted with ours : those in Mosier's vary from 0·14 to 0·64, and in ours from 0·00 to 0·78. In point of fact, the primary intention of Thurstone's Inventory was to measure only a single trait, namely, ' neuroticism,' whereas Bernreuter, as we have seen, deliberately set out to measure several.

A number of other investigators have found a general factor which they identify with neuroticism (see, for example (9), pp. 346f., (10) p. 257, and (19)). But, as Burt points out, " a factor common to all neurotics or to all the traits supposed to be characteristic of neuroticism, is not proved to be a specifically neurotic factor, until it is also shown that the factor in question does *not* appear in the non-neurotic." Thus, in his own investigation, where both a neurotic group and a non-neurotic control group were employed, he finds virtually the same general factor common to both, and identifies it with ' general emotionality ' (7). Thurstone's inventory was originally validated by application to normal students : no external criterion was used, and the only criterion for the validation was that of ' internal consistency.' Nor was any attempt made to discover whether, for instance, students who scored high in the inventory had in fact neurotic tendencies. Hence there would seem to be little direct evidence to justify the assumption that the inventory really measures ' neuroticism.' Mosier's subjects likewise consisted of a group of ordinary students. Many of the symptoms most characteristic of neurotic disturbances (i.e., of such psychoneuroses as hysteria, neurasthenia, hypochondriasis, compulsive and paranoid states) find no place either in his list or in Bernreuter's. The main emphasis is on shyness, sensitiveness, worry, and self-conscious inferiority—in short, as we have indicated, on ' nervousness ' in the popular sense rather than on ' neuroticism ' in the strict sense.

We have already noted that, in our own investigation, nearly all the traits having high saturations for the general factor are to be found in Bernreuter's scale for Introversion (which, as we have seen, largely overlaps with that for Neuroticism).[1] This conscious or unconscious identification of the two seems to rest on a misinterpretation of both terms. First of all, contrary to the popular impression, certain types of functional nervous disorders are by no means uncommon in persons who by temperament are the very opposite of the so-called ' nervous type ' ; hysterics and compulsive-neurotics, for example, usually show extraverted rather than introverted characteristics. It would thus be a gross mistake to confine neuroticism solely to disorders marked by extreme anxiety or unsociability. Secondly, introversion by no means connotes a morbid tendency. Jung, commenting on the misuse of his term, suggests that, being extraverts, both the empirical psychologist and the practical psychiatrist have " a natural bias against the nature of the introvert," and consider his attitude neurotic, and their own quite free from neurotic possibilities. " It is," he adds, " a common mistake to suppose that introversion is more or less identical with neurosis : as concepts the two have no connexion."[2]

Factor II. The first bipolar factor contributes 11·7 per cent. of the total variance, that is, about 25 per cent. of the common factor variance. The largest positive saturations (over ·40) are obtained for the following items : 118. *Keeps in background at social functions.* 94. *Does not take lead at parties.* 121. *Does not like to be with others.* 103. *Has difficulty in talking to strangers.* 87. *Hesitates to introduce people.* 96. *Reluctant to meet important people.* 125. *Self-conscious in starting discussion.* The largest negative saturations are for the following : 71. *Experiences pleasant and unpleasant*

[1] Vernon (20) also points out that the Bernreuter measurement for introversion generally correlates to the extent of about 0·90 with that for neuroticism, and remarks on the fact that in other enquiries the alleged measurements for neuroticism are apt to correlate quite as highly with measurements for introversion as they do with each other.

[2] *Psychological Types* (1923), pp. 472f.

moods. 110. *Does not face troubles alone.* 79. *Understands problems best in discussion.* 38. *Finds conversation more helpful than reading.*

This second factor is by no means easy to interpret. On the whole, judging by the larger saturation coefficients, it would appear to contrast (*a*) the tendency to withdraw from responsibility and to feel uncomfortable in *social* situations with (*b*) the tendency to show signs of emotional instability and dependence on others in more *personal* situations. Thus, broadly speaking, the measurements for this factor might be said to distinguish the person who is introverted (in the *popular* sense, that is, the person who does not enjoy social functions and responsibilities) from the person who is somewhat unstable emotionally and who does not like to be alone or to act solely on the basis of his own judgement.

Owing to the method by which the factors have been extracted, the measurements for this second factor will be uncorrelated with those for the first. Moreover, as will be seen from the table, in this case the saturations for the two factors also show little correlation. Of the items having the largest saturations, whether they belong to the positive or to the negative section, some have high saturations for the first factor and others low : so far as the second factor contrasts the ' unsociable ' with the ' emotionally dependent ' person, it does so regardless of whether they are ' nervous ' or not.

Our second factor in some ways corresponds with the second of Flanagan's two factors. In his analysis this factor " represents about 18 per cent. of the non-chance variance, and is seen to have a large positive weight for Self-sufficiency (·65), a smaller positive weight in Introversion-extraversion (·32), and slight positive weights for Neurotic tendency (·23) and Dominance (·11)." However, as he observes, the contribution of this last is almost negligible. Hence " a high score on this factor would indicate an individual who was independent, but not necessarily dominant " ; at the same time it " indicates nothing with respect to the first factor " (neuroticism). Accordingly he suggests that " Factor Two may be called *sociability*," by which he evidently means *lack* of sociability ((11) pp. 43, 48).

The positive section of our first bipolar factor seems also to show some resemblance to the factor of ' social introversion ' (' *S* ') found in the Guilfords' investigation (12), and to the negative section of the first bipolar factor in Mosier's centroid analysis ((16) Table II). Mosier's factor, when rotated ((16) Table V), yields three or four group factors or ' primary traits,' as he terms them, in which the chief characteristics are ' shyness ' (*S*), ' lack of self-confidence ' (*H* and *P*),[1] and ' sense of inferiority ' (*I*) respectively. In Mosier's analysis the different situations in which these characteristics are reported differ somewhat from those which elicit signs of ' unsociability ' in Bernreuter's list. Some of the situations are more formal, such as public speaking, and others are of a more personal kind, such as being in the presence of ' superiors.' Nevertheless, a tendency to dislike, and to withdraw from, social obligations seems to characterize all the groups mentioned (*S*, *P*, *H*, and *I*) just as it characterizes the items in the positive half of our factor II. And a classification based more on differences between situations than on differences between psychological traits seems clearly discernible in all three sets of results. In our own analysis it should be noted that inferiority has virtually a zero saturation for this factor. Thus Mosier's conclusion that social introversion is closely related to inferiority does not seem to be confirmed. The items in the negative section of our first bipolar factor largely correspond with the items that receive positive signs in Mosier's first centroid factor ((16) Table II). These he eventually subdivided into several component groups—*I* (inferiority), *H* (hypersensitivity), *C* (cycloid, i.e., moodiness, etc.), *D* (depressive, e.g., feeling miserable and worried), *Co* (obsessional,[2] e.g., bothered by useless thoughts), and *Aut* (autistic tendencies, e.g., daydreaming, loneliness). Bernreuter's questions about facing private troubles alone do not appear in Mosier's selection (16). The Guilfords' factor of ' emotional immaturity or dependence ' (*E*) also shows some similarities with this half of our dichotomy, though their questions differ much more from ours ((12), cf. also (9) p. 356).

Factor III. The second bipolar factor contributes 8·4 per cent. of the total variance and 18·7 per cent. of the common factor variance. It introduces a cross-classification of the items as grouped by factor II. In the main it opposes the items in what Bernreuter describes the Neurotic and the Introverted sections of his inventory to those which are intended to reveal Dominance and Self-sufficiency. As the traits are arranged in Table II (i.e., with certain of Bernreuter's designations reversed [3])

[1] ' *H* ' denotes ' hypersensitivity ' ; ' *P* ' lack of confidence in public speaking.

[2] Mosier hesitates between this description and one which would attribute this factor to some defect of cognition or attention.

[3] It will be remembered that two of the items in the scale B2–S were not reversed. In this factor these are classed with the items in scales B3–I and B1–N.

the positive portion of factor III indicates a lack of dominance and of self-sufficiency and the negative portion a lack of stability or a tendency to introversion in Jung's original sense. Thus within the positive portion of the first bipolar (the portion dealing with social situations) this factor contrasts (1*a*) those who cannot control *others* in a social situation (lack of ' dominance ') with (1*b*) those who cannot control *their own feelings* in such a situation (lack of stability). Within the negative portion (that dealing with personal situations) it contrasts (2*a*) those who are dependent on the advice of *others* and (2*b*) those who are at the mercy of *their own feelings or moods.*

The four groups also exhibit a rough parallelism with a fourfold classification which emerged from a bipolar analysis of Cattell's correlation table, carried out by one of us and published in a recent number of this *Journal.* The correlations were based, not on self-ratings, but on assessments supplied by observers for 240 college students. There the first bipolar factor divided the several types into ' dysthymic ' and ' euthymic.' Bernreuter's items, with the reversals used for the present analysis, relate, as we have seen, almost exclusively to what are popularly termed ' nervous ' characteristics, and include little or nothing which would class a man as positively ' euthymic.' Hence the four groups found with his scales correspond most directly with the four subsections of the ' dysthymic ' group. Thus (1*a*) in the Bernreuter series—' lack of dominance ' (ability to control others in social situations) corresponds with IV A in the dysthymic group and with the negative of IV E (dominance) in the euthymic ; (1*b*) ' lack of ability to control one's own feelings in social situations ' corresponds with IV C (anxiety) in the dysthymic group and the negative of IV G (sociability or cordiality) in the euthymic group ; (2*a*) ' lack of self-sufficiency ' (i.e., of independence of others in personal situations) corresponds with IV D (immaturity) in the dysthymic group and the negative of IV H (facile) in the euthymic ; (2*b*) ' lack of independence of feelings in personal situations ' corresponds with IV B (' compulsive ') in the dysthymic group and the negative of IV F (' impulsive ') in the euthymic, although several of the traits overlap with IV G and the negative of IV G ((1) p. 210, Table II).[1]

On the whole, it will be seen, the subdivisions indicated by our third factor are consistent with those originally proposed by Bernreuter, although, since so many questions are common both to the Neurotic and to the Introverted section, his classification tends to become threefold rather than fourfold. Since his inventory is concerned not so much with personality in general as with those aspects of personality which are distinctive of the nervous or ' neurotic ' personality, it is not easy to relate these factors with the results of those analyses that attempt to deal with the whole range of a man's affective constitution (for these, cf. (9) pp. 538 and 550). Certainly, the factors of instability and of extraversion have appeared in many earlier studies, even as far back as Burt's 1915 study.[2] Others are somewhat suggestive of some of McDougall's instinctive propensities (which in factorial studies appear as *groups* of tendencies), e.g., dominance or ' assertiveness,' social introversion or lack of ' sociability,' lack of self-sufficiency or ' submissiveness.' On the other hand, Bernreuter's list contains very few questions relating to some of the most important factors that emerge in investigations on normal persons (e.g., the factor of euphoria as contrasted with dysphoria) ; and it makes very little mention of the symptoms associated with such neurotic disorders as hysteria, neurasthenia, compulsion-neurosis, paranoid and hypochondriacal tendencies, and the like. Flanagan's statement that the inventory " includes most of the psychological traits other than ' intelligence ' which recent investigators have attempted to measure " seems decidedly misleading.[3]

General Conclusions. Several broader inferences may be drawn from a survey of the results furnished by the foregoing analysis. First, both the original correlations and the factor analysis indicate that, as judged by their most diagnostic items, Bernreuter's scales for Dominance and Self-sufficiency are fairly consistent within themselves : in the third factor these two groups of items are related to each other, statistically as well as logically, but in the second factor they are negatively correlated ; in the first and second factors they show some slight relation with the items in the Neurotic and Introvert scales, but are negatively correlated with them in the third factor. Secondly, the items introduced into the Neurotic and Introvert scales overlap to a very large extent with each other, but, taken as a whole, they

[1] Those who are interested in Cattell's classification will find that the foregoing have some affinity with the four factors which Cattell, in his final list, has labelled IV or E (Submissiveness, including QP VI), VI or I (Anxious Emotionality, including QP VII), VIII or G (Dependent Character, including QP IIa and VII), and XII or L (Surgent Cyclothymia, including QP V) ((9) pp. 356f. and 577f.).

[2] ' General and Specific Factors Underlying the Primary Emotions,' *loc. cit. sup.*

[3] It is certainly incorrect if ' investigators ' is intended to include British investigators working in the field at about that date : (cf., for example, Burt, *The Young Delinquent,* 1925, pp. 399f., and (6) pp. 353f.).

appear to be reasonably consistent : in the general factor they make a very appreciable contribution to the total variance. And thirdly, we think our results make it clear that factor analysis has still a large and useful part to play in the classifying of items in personality inventories. By a more judicious choice of items, based on factorial results, it should be possible to construct both a broad and well-balanced inventory designed to measure the general pattern of ' personality ' in any normal individual, and a narrower and more intensive inventory intended to assess the individual's liability to develop psychoneurotic disorders. But as has frequently been pointed out, " American questionnaires and American results are applicable only to Americans." Hence there is an urgent need for standard schemes for British use.

In this country questionnaires of the former type, usually based on McDougall's scheme of instincts and sentiments (cf. (8) and refs.) are to be found in several unpublished theses [1] ; of the latter type the best known is probably the ' Questionnaire on Neurotic Symptoms ' drawn up by Burt (6).

The practical difficulties presented by most of the existing inventories seem to be twofold : first, their length, which in many cases is out of all proportion to the information obtained ; and secondly, their lack of balance : important traits and symptoms are often omitted altogether ; on the other hand, many of the questions are plainly repetitions of the same familiar theme, and others are shown by the statistical analysis to overlap in ways not obvious to inspection. To correct these two main defects, factorial methods would in our view [2] prove most helpful. Both bipolar and group factor analyses should be tried. Writers who incline to hold that neurotic disorders form distinct clinical entities will probably, with Mosier, prefer an analysis into group factors. But for more practical purposes a bipolar analysis leading to a progressive classification of a hierarchical type would seem more suitable. That the more significant symptoms do fit into such a logical structure is indicated by our own results, and, perhaps even more clearly, by those printed by Mosier. Although he does not call attention to the fact, his eight group factors, secured by rotation, coincide almost exactly with the eightfold divisions of his fourth centroid factor ; and the preceding factors in his centroid table indicate broader and broader classifications of these narrower groups in accordance with the requirements of hierarchical structure.

The most unsatisfying feature about the lists of factors put forward by previous factorists (e.g., those enumerated by Cattell) is their miscellaneous and incoherent character. It might, of course, turn out to be an empirical fact that, on the orectic side if not on the cognitive, the mind does consist merely of a chaotic set of traits. Nevertheless with an appropriate factorial procedure, such as will allow a hierarchical structure to emerge when present, a more systematic pattern is nearly always discernible. When such a pattern does appear, it will, we think, be almost universally admitted that it renders the interpretation offered far more plausible. As an example, we need only refer to the study already quoted, where a set of miscellaneous factors was found, on re-analysis, to fit readily into a hierarchical scheme (1).[3]

But, as in all factorial investigations, it is essential to check conclusions derived in the first instance from internal evidence only by comparisons with criteria based on external evidence. In particular, the inferences drawn from self-ratings and self-inventories need to be checked by data obtained from ratings of the same individuals supplied by competent external observers or better still derived from observations of actual behaviour.

[1] Of these perhaps the most thorough is that drawn up by A. R. Bassam in a research carried out at University College (' A Factorial Study of Emotional Assessments,' Ph.D. Thesis, 1945). His investigation is one of the very few in which the same individuals have been assessed both by themselves and by other observers.

[2] Cattell also remarks how " the natural unreliability of self-ratings by questionnaires is still further exaggerated by the lack of critical insight in the designers themselves when setting up their ' unitary traits ' : obviously these categories of the market place cannot be accepted as true functional unities ' ((9) p. 347).

He himself, however, in his brief review of work based on the " meretricious self-inventory," is far more pessimistic about the value of factorial investigations in this field. Yet the majority of his criticisms point, not so much to the worthlessness of such questionnaires, but rather to the need for improving them in the light of more objective evidence and of hypotheses based directly on the psychology of the normal personality instead of on the conceptions carried over from the abnormal to the normal field.

[3] As a result of his studies of neurotic patients, Eysenck also concludes that " a hierarchical structure of personality, akin to Burt's conception of factors and McDougall's theory of sentiments, would appear to account for the factor better than any other " ((10) p. 16).

III. SUMMARY

1. The Bernreuter Personality Inventory was applied to 100 students of psychology, and tetrachoric correlations were calculated between the 32 items selected by Bernreuter as the ' most diagnostic.' The correlation table thus obtained was then factorized by Burt's Method of Simple Summation, with reduced self-correlations. Three significant non-correlated factors were found.

2. The first factor, contributing about 23 per cent. to the total variance, appears to be a general factor of nervousness in the popular sense. The second factor, contributing about 12 per cent. to the total variance, appears to contrast the person who dislikes social situations with the person who shows signs of emotional instability and dependence on others in private situations. The third factor, contributing about 8 per cent., contrasts dependence on others with dependence on one's own moods and feelings.

3. These factors have been compared, as regards content, with those obtained by other investigators with somewhat similar questionnaires. They show the closest resemblance with the centroid or bipolar factors obtained by Mosier (before rotation) from his analysis of the most diagnostic questions in Thurstone's ' Personality Inventory.' On the other hand, the analysis demonstrates very clearly that, as they stand, the questions analysed do not cover the whole of normal personality, nor do they suffice to indicate all the commoner forms of neurotic tendency.

REFERENCES

1. Banks, C. (1948). ' Primary personality factors : a re-analysis.' *Brit. J. Psych., Stat. Sect.*, I, 204–18.
2. Bernreuter, R. B. (1933). ' The theory and construction of the Personality Inventory.' *J. Soc. Psychol.*, IV, 387–405.
3. Bernreuter, R. B. (1933). ' The validity of the Personality Inventory.' *Person. J.*, XI, 383–6.
4. Bernreuter, R. B. (1935). *Manual for the Personality Inventory*. Stanford : The Stanford University Press.
5. Buros, O. K. (1941). *The Nineteen Forty Mental Measurements Yearbook*. New Jersey : Mental Measurements Yearbook, Highland Park.
6. Burt, C. (1935). *The Subnormal Mind*. London : Oxford University Press.
7. Burt, C. (1948). ' The factorial study of temperamental traits.' *Brit. J. Psych., Stat. Sect.*, I, 178–203.
8. Burt, C. (1950). ' The factorial study of emotions,' *ap.* Reymert, M. L. (ed.). *Feelings and Emotions*. New York : McGraw-Hill Book Co.
9. Cattell, R. B. (1946). *Description and Measurement of Personality*. New York : World Book Company.
10. Eysenck, H. J. (1947). *Dimensions of Personality*. London : Kegan Paul, Trench, Trubner, and Company.
11. Flanagan, J. C. (1935). *Factor Analysis in the Study of Personality*. Stanford : The Stanford University Press.
12. Guilford, J. P., and Guilford, R. B. (1936). ' Personality factors, S, E, and M, and their measurement.' *J. Psychol.*, II, 109–28.
13. Landis, C., and Katz, S. E. (1934). ' The validity of questions which purport to measure neurotic tendencies.' *J. Appl. Psychol.*, XVIII, 343–56.
14. Landis, C., Katz, S. E., and Zubin, J. (1935). ' Empirical evaluation of three personality adjustment inventories.' *J. Educ. Psychol.*, XXVI, 321–30.
15. Lorge, Irving (1935). ' Personality traits by fiat. I.' *J. Educ. Psychol.*, XXVI, 273–8.
16. Mosier, C. I. (1937). ' A factor analysis of certain neurotic tendencies.' *Psychometrika*, II, 236–87.
17. Sartain, A. Q. (1946). ' A study of Bernreuter Personality Inventory scores.' *J. Soc. Psychol.*, XXIV, 255–9.
18. Stagner, R. (1934). ' Validity and reliability of the Bernreuter Personality Inventory.' *J. Abn. Soc. Psychol.*, XXVIII, 413–18.
19. Super, D. E. (1942). ' The Bernreuter Personality Inventory : a review of research.' *Psychol. Bull.*, XXXIX, 94–125.
20. Vernon, P. E. (1938). ' The assessment of psychological qualities by verbal methods.' *Industrial Health Research Board Reports, No.* 83. London : H.M. Stationery Office.

APPENDIX

TABLE III. ITEMS USED IN THIS STUDY

Number	Question	Diagnostic Response	Frequency of Responses Yes	No	?
	B1–N				
5*	Can you stand criticism without feeling hurt?	No	67	19	14
8*	Do you often feel just miserable?	Yes	26	66	8
24*	Are you troubled with shyness?	Yes	28	65	7
39	Do you worry too long over humiliating experiences?	Yes	32	60	8
51*	Are your feelings easily hurt?	Yes	34	58	8
54	Do you often feel lonesome when you are with other people?	Yes	26	68	6
72*	Are you troubled with feelings of inferiority?	Yes	32	54	14
73	Does some particularly useless thought keep coming to your mind to bother you?	Yes	14	82	4
104*	Do your feelings alternate between happiness and sadness without apparent reason?	Yes	15	81	4
106*	Do you worry over possible misfortunes?	Yes	32	63	5
	B2–S				
23	Do you think you could become so absorbed in creative work that you would not notice a lack of intimate friends	Yes	33	62	5
38	Do you usually find conversation more helpful in formulating your ideas than reading?	No	50	30	20
61	Do you usually enjoy spending an evening alone?	Yes	47	37	16
79	Can you usually understand a problem better by studying it out alone than by discussing it with others?	Yes	42	40	18
82	Are you willing to take a chance alone in a situation of doubtful outcome?	Yes	87	7	6
89	Do you like to get many views from others before making an important decision?	No	64	30	6
101	Do you usually prefer to work with others?	No	47	34	19
110	Do you usually face your troubles alone without seeking help?	Yes	57	33	10
112	Do you prefer to make hurried decisions alone?	Yes	59	27	14
121	Do you like to be with other people a good deal?	No	53	21	26
	B3–I				
2	Do you daydream frequently?	Yes	39	54	7
5*	Can you stand criticism without feeling hurt?	No	67	19	14
8*	Do you often feel just miserable?	Yes	26	66	8
12	Do you blush very often?	Yes	22	73	5
14	Do you consider yourself a rather nervous person?	Yes	22	73	5
39	Do you worry too long over humiliating experiences?	Yes	32	60	8
51*	Are your feelings easily hurt?	Yes	34	58	8
71	Do you experience many pleasant and unpleasant moods?	Yes	46	38	16
104*	Do your feelings alternate between happiness and sadness without apparent reason?	Yes	15	81	4
106*	Do you worry over possible misfortunes?	Yes	32	63	5
	B4–D				
24*	Are you troubled with shyness?	Yes	28	65	7
40	Have you ever organized any clubs, teams, or other groups on your own initiative?	Yes	68	30	2
72*	Are you troubled with feelings of inferiority?	No	32	54	14
80	Do you lack self-confidence?	No	68	30	2
87	Do you take the responsibility of introducing people at a party?	Yes	54	27	19
94	Do you ever take the lead to enliven a dull party?	Yes	52	34	14
96	At a reception or tea do you feel reluctant to meet the most important person present?	No	22	68	10
103	Do you have difficulty in starting a conversation with a stranger?	No	27	68	5
118	Do you keep in the background at social functions?	No	38	44	18
125	Would you feel very self-conscious if you had to volunteer an idea to start a discussion among a group of people?	No	15	79	6

*These items, it will be noted, occur in *two* of Bernreuter's sections or 'scales.'

Vol. V Part I . The British Journal of Psychology, Statistical Section March, 1952

THE APPLICATION OF ANALYSIS OF VARIANCE TO PROBLEMS OF CORRELATION

By A. E. G. PILLINER
Moray House, University of Edinburgh

I. *Previous Investigations.* II. *The Case of Two Arrays.* III. *A Comparison of the Two Assessments of the Correlation.* IV. *The General Case of* n *Arrays using Variance Components.* V. *An Extension to Three-way Analysis.* VI. *Summary and Conclusions.*

I. PREVIOUS INVESTIGATIONS

The possibility of assessing correlation coefficients by means of the technique of analysis of variance has already been noted by several earlier writers. Burt has discussed the problem in *The Factors of the Mind* ((1) Chap. X) ; and in a later article (2) he has used the procedure to estimate the reliability of teachers' assessments of their pupils, enumerating other applications in an appendix. Hoyt (6) has employed the method to assess the reliability of a test where all the items are of a pass-fail type ; he arrives at formulæ identical with those already derived by Kuder and Richardson (7).

The present article deals with certain theoretical implications of the technique, and suggests extensions in which components of variance are used (cf. Snedecor (8), Chaps. X, XI ; Tippett (9), Chaps. VI, X). It will be assumed that the array variances in the population are equal. The numerical illustrations are based on data employed by Finlayson in an enquiry on the marking of essays (5) : it was indeed for this enquiry that the procedures to be described were originally worked out.

II. THE CASE OF TWO ARRAYS

Consider first the case of two tests, each administered to the same group of N children. If an analysis of variance is performed on the paired scores, the resulting sums of squares will be as shown in Table I.

TABLE I. SUMS OF SQUARES WITH TWO TESTS AND N CHILDREN

Source	Sum of Squares
Tests	$\frac{1}{2}N\bar{d}_{12}^2$ $= S_T$ say
Children	$H_1 + H_2 - \frac{1}{2}D_{12} = S_C$ say
Error	$\frac{1}{2}D_{12}$ $= S_E$ say
Total	$H_1 + H_2 + \frac{1}{2}N\bar{d}_{12}^2$

Here H_1 and H_2 are the respective sums of squares 'within tests,' D_{12} those of the differences between paired scores all measured from their means, while $\bar{d}_{12}$ is the difference between the test means. The product-moment correlation between the tests is [1]

$$r_{12} = \frac{H_1 + H_2 - D_{12}}{2\sqrt{(H_1 H_2)}} . \qquad (1)$$

[1] This is the familiar 'difference formula for r' : cf., for example, Kelley, T. L., *Statistical Method* (1923), p. 180, eq. 130, and (2) p. 92.

Now, when $H_1 = H_2$, $2\sqrt{(H_1H_2)} = H_1 + H_2$; and when $H_1 \neq H_2$, $2\sqrt{(H_1H_2)} < H_1 + H_2$. Hence, writing

$$r'_{12} = \frac{H_1 + H_2 - D_{12}}{H_1 + H_2}, \qquad (2)$$

we find $r_{12} = r'_{12}$ if $H_1 = H_2$; $r_{12} > r'_{12}$ if $H_1 \neq H_2$. From (2) and Table I we see [1] that

$$r'_{12} = \frac{S_C - S_E}{S_C + S_E}. \qquad (3)$$

Thus an assessment of the correlation between the two tests is obtainable from the analysis of variance. This assessment is unbiased when the observed test variances are the same; and in any case the coefficient of correlation computed by the product moment method will never be less than the value obtained as above from an analysis of variance.

III. A COMPARISON OF THE TWO ASSESSMENTS OF THE CORRELATION

It will be of interest to compare more precisely the values for r_{12} and r'_{12} as derived from an analysis of variance in place of r_{12}. The latter is converted into the former by multiplying by the factor $(H_1 + H_2)/2\sqrt{(H_1H_2)}$. This follows from eq. (1) and (2). If we write $a = H_2/H_1$ then $(H_1 + H_2)/2\sqrt{(H_1H_2)} = (1 + a)/2\sqrt{a}$, and

$$r_{12} = r'_{12}(1 + a)/2\sqrt{a}. \qquad (4)$$

It is helpful to construct a table setting out the values of the multiplier shown in eq. (4), corresponding to given values of a. Such a table shows that the discrepancy is quite small, particularly for small values of a. Even when the ratio of the variances is as high as 1·5, the discrepancy between r_{12} and r'_{12} is only about 2 per cent. When the ratio of the variances reaches 4, r'_{12} is about 80 per cent. of r_{12}.

Other considerations, however, enter here. While the technique described can be employed to obtain the product moment correlation between two variables, it would be pointless to use it if the correlation were in fact the only end desired. It is only when other questions are to be asked, which can best be answered by analysis of variance methods, and when the correlation is required as well, that this technique is proposed. This limits its use to cases when analysis of variance technique can be legitimately applied, that is, when the within-test variances (and hence H_1 and H_2) do not differ significantly. For example, if the data were intelligence quotients from a pair of parallel tests, the hypothesis might be adopted that the population variances are equal, the discrepancies being due to sampling error. With this hypothesis, their mean would provide a better estimate of the within-test sum of squares than either separately. If this were so, the denominator of eq. (1) would be replaced by $(H_1 + H_2)$. Thus r'_{12} would be a better estimate of ρ_{12}, the population value, than r_{12}. The decision whether to adopt such a hypothesis must depend on circumstances.

IV. THE GENERAL CASE OF n ARRAYS USING VARIANCE COMPONENTS

If there are more than two tests, more general equations may be derived. Suppose an examination consisting of a battery of n parallel tests is administered to N children. The score of the ith child for the jth test may be written :—

$$x_{ij} = c_i + t_j + w_{ij}, \quad (i = 1, 2, \ldots, N;\ j = 1, 2, \ldots, n),$$

[1] This expression for the correlation coefficient is given by Burt (Report of Internat. Inst. Exam. Enquiry on *Marks of Examiners*, p. 255) who interprets it factorially by reference to the axes of the correlation ellipse. Cf. also (1) p. 275, eq. i, putting k, the number of variables correlated, = 2.

where the terms denote independent contributions to this score specific to child, test, and interaction (here used as error), respectively. Hence the child's total score for all the tests will be

$X_i = \sum_{j=1}^{n} x_{ij} = nc_i + \sum_{j=1}^{n} t_j + \sum_{j=1}^{n} w_{ij}$. Since the second of the terms on the right is common to all children, the variance of the child totals X_i ($i = 1, 2, \ldots, N$) is $\sigma^2_X = n^2\sigma^2_C + n\sigma^2_E$, where σ^2_C and σ^2_E are the population variances for children and for error respectively ((8) p. 261, (9) p. 215).

If the whole examination is now repeated, the portion of the total variance common to both administrations is $n^2\sigma^2_C$. Hence the reliability of the examination will be

$$\rho_{nn} = n\sigma^2_C/(n\sigma^2_C + \sigma^2_E). \qquad (5)$$

If there is only one test in each examination, i.e., $n = 1$, this reduces to the familiar equation ((2) p. 82)

$$\rho_{11} = \sigma^2_C/(\sigma^2_C + \sigma^2_E). \qquad (6)$$

The relation between (5) and (6) is made clear if the Spearman-Brown formula is applied to ρ_{11}; the result is ρ_{nn}. Thus ρ_{11} is the average inter-correlation of two tests derived from the battery of n tests.

Estimates of σ^2_C and σ^2_E are obtainable from a two-way table of analysis of variance (tests, children, error). If S_C and S_E are the sums of squares in the analysis of variance, and m_C and m_E the mean squares for children and error, $S_C/(N-1) = m_C \rightarrow n\sigma^2_C + \sigma^2_E$, and $S_E/(N-1)(n-1) = m_E \rightarrow \sigma^2_E$. Hence it is reasonable to take as an estimate of ρ_{nn}

$$r'_{nn} = (m_C - m_E)/m_C\ ; \qquad (7)$$

and as an estimate of ρ_{11},

$$r'_{11} = (m_C - m_E)/\{m_C + (n-1)m_E\} = \{(n-1)S_C - S_E\}/(n-1)(S_C + S_E) \qquad (8)$$

(cf. Burt (2), p. 90, eq. (v), and (1), p. 275, eq. (ii), for 'unbiased r'). When $n = 2$, eq. (8) reduces to eq. (3). It follows, as Burt points out, that estimates of the reliability of an examination, and of the mean inter-correlation between the tests in the examination, can readily be obtained by means of an analysis of variance.

The assumption that the test variances in the population are equal should be noted. In the case of exact equality the estimates are unbiased; when the variances are not equal, the estimates are, it would seem, negatively biased. Here again, if the tests are parallel, it may be argued that the observed deviations from equality of the sample variances are random, and that the best estimate of the population variance is their mean. With this hypothesis the estimates of the mean inter-correlation and reliability from the analysis of variance are valid.

Table II illustrates the use of these equations. The data are the marks assigned independently by six examiners (marking by general impression) to single essays written by twenty children, a random sample from a larger group investigated by Finlayson (5).

TABLE II. ANALYSIS OF VARIANCE OF MARKS ASSIGNED TO SINGLE ESSAYS

Source	d.f.	Sums of Squares	Mean Square	Components
Markers ..	5	294·07	58·81	$N\sigma^2_M + \sigma^2_E$
Children ..	19	763·14	40·16	$M\sigma^2_C + \sigma^2_E$
Error.. ..	95	254·26	2·68	σ^2_E
Total ..	119	1311·47	—	—

M = number of markers (6); N = number of children (20).

From eq. (8), and noting that tests have become markers (i.e., persons), we find the estimated mean inter-correlation to be ·7001. Similarly, from eq. (7), the reliability of the whole examination is found to be ·9334. To estimate the reliability for intermediate numbers of markers, we proceed as follows. The estimated variance components (see column on right of table) are $2{\cdot}807 \rightarrow \sigma^2_M$; $6{\cdot}248 \rightarrow \sigma^2_C$; $2{\cdot}676 \rightarrow \sigma^2_E$. Using eq. (5), and again noting that tests have become markers, we can estimate the reliability of the whole examination for different numbers of markers.

TABLE III. ESTIMATED MARKER RELIABILITIES

Markers	r'_{MM}	Markers	r'_{MM}
$M = 1$	·7001	$M = 4$	·9033
$M = 2$	·8236	$M = 5$	·9211
$M = 3$	·8751	$M = 6$	·9334

If for any reason it is desired to obtain the mean product-moment inter-correlation, the method can still be used, though less directly than with two tests only. Consider the case where all the within-test variances are the same, and denoted by σ^2_W. We may then drop the prime and write $r_{11} = \{(n-1)\,S_C - S_E\}/(n-1)\,(S_C + S_E)$. Now $S_C + S_E = S_W$, the sum of squares within tests. Hence $r_{11} = (n\,S_C - S_W)/(n-1)\,S_W$. Further, under these conditions, $S_C \rightarrow \dfrac{N-1}{n}\sigma^2_X$, so that

$$r_{11} \rightarrow \frac{\sigma^2_X - n\sigma^2_W}{n\,(n-1)\sigma^2_W} = \frac{1}{n-1}\left\{\frac{\sum_{i=1}^{N}{}' X^2_i}{n\sum_{i=1}^{N}{}' x^2_i} - 1\right\}. \qquad (9)$$

(The prime after the summation sign denotes that scores are measured from their mean.)

In practice the test variance will not often be the same in each test; but a simple scaling procedure will make it so without disturbing the inter-correlations. Compute first the several within-test sums of squares, S_j. With the largest (which we may call S) as numerator in each case, find the ratios S/S_j and calculate their square roots. Use these as factors to weight the scores in each test appropriately and from the weighted scores compute for each test a new sum of squares; the results for all tests should be the same and their sum gives $n\sum_{i=1}^{N}{}' x^2_i$. Obtain the values X_i by summing for children *across* the tests after weighting and hence calculate $\sum_{i=1}^{N}{}' X^2_i$. Finally apply eq. (9) to obtain the true mean inter-correlation. The method is similar to that employed by Edgerton and Toops (4), but is probably simpler to apply.

Using the same data as for Table II we find for the true mean inter-correlation ·7272 which should be compared with ·7001 from the analysis of variance. The difference is the result of different sums of squares for markers.

At Burt's suggestion Miss Cast also investigated the marking of compositions (3); and it may be of interest to compare her results.[1] In Table IV of her article (Part I) she uses analysis of variance to compare different marking systems, and in Table VII (Part II) the mean product-moment inter-correlations are quoted. The estimated inter-correlations calculated from the analysis of variance are given below in Table IV, together with the coefficients calculated by the product-moment formula. As with Finlayson's data, it will be seen that r_{11} is always greater than r'_{11}.

[1] A somewhat similar use of an analysis of variance was also adopted in the later Report of the Internat. Inst. Exam. Enquiry on *The Marking of English Essays* (1941), pp. 46f.: cf. p. 149.

TABLE IV. MEAN INTER-CORRELATIONS CALCULATED BY BOTH METHODS (MISS CAST'S DATA)

Method	r'_{11}	r_{11}
Individual	·431	·479
Achievement	·423	·439
General impression	·465	·485
Analytic	·473	·492

To return to eq. (8), as Burt has pointed out, this is of exactly the same form as that for the intra-class correlation (cf. (1), p. 275, footnote). Moreover, the method of deriving it is similar (cf. Snedecor (8), p. 243). The difference appears to reside in the number of degrees of freedom appropriate to the two cases. Even if the marker means are identical, so that the sums of squares for children and error are respectively the same whether one-way or two-way analysis is adopted, the number of degrees of freedom for error will be less in the latter case than in the former by the number of degrees of freedom for markers.

V. AN EXTENSION TO THREE-WAY ANALYSIS

This is not the complete story, however. Not only is the marking not completely reliable, but also the children may be inconsistent in their performances. This can be investigated by arranging for *two* essays to be written by each child, and each essay to be marked by the same group of markers.

The principle is the same as in the previous simpler case. Suppose an examination consisting of n different essays is given to N children, each essay being marked independently by the same M markers, that is, each child will have Mn separate scores on the whole examination.

Let us denote the score of the ith child for the jth essay marked by the kth marker by

$$x_{ijk} = c_i + t_j + m_k + u_{ij} + v_{ik} + w_{jk} + z_{ijk}$$
$$(i = 1, 2, \ldots, N\,;\; j = 1, 2, \ldots, n\,;\; k = 1, 2, \ldots, M),$$

where the terms respectively stand for independent contributions to this score specific to child, test, marker, three first-order interactions, and one second-order interaction (the last being considered as error).

The child's total score on the n essays, each marked by M markers, will be

$$X_i = MnC_i + M\sum_{j=1}^{n} t_j + n\sum_{k=1}^{M} m_k + M\sum_{j=1}^{n} u_{ij} + n\sum_{k=1}^{M} v_{ik} + \sum_{j=1}^{n}\sum_{k=1}^{M} w_{jk} + \sum_{j=1}^{n}\sum_{k=1}^{M} z_{ijk}.$$

Of the terms in the expression on the right, only the first, fourth, fifth, and seventh will alter from child to child, so that the variance of child totals X_i $(i = 1, 2, \ldots, N)$ is

$$\sigma^2_X = M^2n^2\sigma^2_C + M^2n\,\sigma^2_{CE} + Mn\sigma^2_{CM} + Mn\sigma^2_{CEM},$$

where the suffixes refer respectively to child, child-essay and child-marker interactions, and second-order interaction. All these components can be estimated from the analysis of variance.

Coming to correlations, there are now four possible reliability coefficients.

(1) If the whole examination is repeated, that is, the same N children write essays on the same n topics as before, and each essay is marked by the same M markers as before, the first three terms of the above variance expressions will be common to both applications and the reliability of the examinations will be

$$\rho_1 = \frac{Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM}}{Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM} + \sigma^2_{CEM}}. \qquad (10)$$

(2) If the conditions are as above, except that a different set of markers is concerned with the second examination, only the first and second terms are common to both examinations and the reliability will be

$$\rho_2 = \frac{Mn\sigma^2_C + M\sigma^2_{CE}}{Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM} + \sigma^2_{CEM}}. \qquad (11)$$

(3) If on the other hand the essays differ on the two occasions and the markers remain constant, the first and third terms are common and the reliability will be

$$\rho_3 = \frac{Mn\sigma^2_C + n\sigma^2_{CM}}{Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM} + \sigma^2_{CEM}}. \qquad (12)$$

(4) Finally, if both essays and markers differ on the two occasions, the only common term will be the first; and the reliability will be

$$\rho_4 = \frac{Mn\sigma^2_C}{Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM} + \sigma^2_{CEM}}. \qquad (13)$$

Once the variance components have been estimated, the estimates of the above four different correlations can be calculated for various values of M and n.

The following problem may serve to illustrate a case in which the foregoing technique might usefully be employed. An education authority, we may suppose, wishes to know what reliability could be placed on the subjective marking of examination scripts under a variety of conditions: a study of the analysis of variance resulting from a pilot experiment, and of the estimated correlations, would then indicate how many scripts ought to be written by each child and how many markers should be engaged in order to achieve a given degree of reliability. Or, to take another example, the notoriously difficult problem of interviewing might be investigated on much the same lines. Here, however, I will again use data from Finlayson's enquiry to illustrate the working procedure.

An analysis of variance was carried out for Finlayson's figures, and the required variance components [1] calculated as shown in Table V: (S.Sq. denotes 'sum of squares' and M.Sq. 'mean square').

TABLE V. ANALYSIS OF VARIANCE OF MARKS ASSIGNED TO TWO ESSAYS PER CHILD

(6 markers, 20 children, 2 essays per child)

Source	d.f.	S.Sq.	M.Sq.	Components
Markers ..	5	426·033	85·21	$Nn\sigma^2_M + N\sigma^2_{ME} + n\sigma^2_{CM} + \sigma^2_{CEM}$
Essays ..	1	0·816	0·82	$MN\sigma^2_E + M\sigma^2_{CE} + N\sigma^2_{ME} + \sigma^2_{CEM}$
Children ..	19	1,118·316	58·86	$Mn\sigma^2_C + M\sigma^2_{CE} + n\sigma^2_{CM} + \sigma^2_{CEM}$
M × E ..	5	31·734	6·35	$N\sigma^2_{ME} + \sigma^2_{CEM}$
M × C ..	95	256·634	2·70	$n\sigma^2_{CM} + \sigma^2_{CEM}$
E × C ..	19	92·184	4·85	$M\sigma^2_{CE} + \sigma^2_{CEM}$
M × E × C	95	182·266	1·92	σ^2_{CEM}
Total ..	239	2,107·983	—	

Solving for the components, we have the estimates

$$4{\cdot}4354 \rightarrow \sigma^2_C\ ;\ {\cdot}4888 \rightarrow \sigma^2_{CE}\ ;\ {\cdot}3910 \rightarrow \sigma^2_{CM}\ ;\ 1{\cdot}9190 \rightarrow \sigma^2_{CEM}.$$

The estimated reliabilities for these components, calculated from eq. (10) to (13), are given below for values of $n = 1$ and 2 and of $M = 1, \ldots, 6$. It should be noted that the Spearman-Brown formula no longer applies except in the case of r_{12}.

[1] For this method of subdividing the variance obtained from children's marks, the reader may refer to this *Journal*, I, p. 10, Table IV.

TABLE VI. ESTIMATED RELIABILITIES

	r'_1		r'_2		r'_3		r'_4	
n	1	2	1	2	1	2	1	2
1	·7347	·8409	·6807	·7760	·6672	·8004	·6131	·7355
2	·8422	·9104	·8100	·8739	·7618	·8648	·7296	·8283
3	·8877	·9377	·8648	·9122	·8018	·8900	·7789	·8646
4	·9128	·9522	·8950	·9327	·8240	·9035	·8062	·8840
5	·9287	·9612	·9142	·9454	·8380	·9119	·8235	·8961
6	·9398	·9674	·9275	·9541	·8477	·9176	·8354	·9043

It is clear from eq. (10) to (13) that r'_1 will always be the largest and r'_4 the smallest of the four possible reliabilities. In the former case, correlation of errors is a maximum, and in the latter a minimum ; the other two cases, r'_2 and r'_3, are intermediate.

It is of interest to compare the correlations in the above table with those obtained by the ordinary product-moment procedure. Thus, if the marks obtained on the two essays are pooled for each child and marker (i.e., each child will have six entries, each the sum of two essay marks), then fifteen marker inter-correlations are possible, with a mean of ·7958. The value corresponding to this in Table VI is ·7760 (r'_2, $M = 1$, $n = 2$). The effect of the differences between the marker variances of the summed scores of two essays is obvious. Even so, although the ratio of the highest and the lowest standard deviations was in this case as high as 1·75, the two estimates are in fair agreement.

Again, the marks awarded by the six markers to each child on the first essay may be pooled and correlated with those awarded by the same six markers to the second essay, the result being ·8781. This is to be compared with the value in Table VI of ·8477 (r'_3, $M = 6$, $n = 1$). In this case we may obtain the product-moment correlation from the estimate by applying the a-correction. It is noteworthy that for these data $a = 1·71$, yet the difference is only about ·03.

A criticism that might perhaps be levelled against the investigation here used by way of illustration is that only two essays were written by each child, so that the estimate of σ^2_{CE} in particular may be somewhat inaccurate ; however, the general principle remains sufficiently clear.

VI. SUMMARY AND CONCLUSIONS

1. It is concluded that, when the hypothesis of equal test variances in the population is satisfied, the best sample estimate of the correlation is that obtained from an analysis of variance, whereas the coefficient obtained from the sample by the product-moment method is an overestimate.

2. In a two-analysis with n arrays ($n > 2$), the components of the variance, obtained by an analysis of variance, provide estimates of the reliability coefficients for different values of n. As Burt has indicated ((1) p. 275), the mean inter-correlation obtained is analogous to, but not identical with, the intra-class correlation. A short method of obtaining the mean product-moment inter-correlation is described.

3. The use of variance components is further extended to the more complex case of three-way analysis ; and various coefficients of reliability are obtained, under stated conditions, from a table of analysis of variance. The majority of these could not have been arrived at by the Spearman-Brown formula.

4. It is contended that such pilot investigations as are here envisaged should be so planned that every factor concerned is replicated as often as possible ; the variance components are then estimated with greater accuracy.[1]

[1] The writer wishes to express his grateful acknowledgements to Dr. D. N. Lawley and to Mr. W. G. Emmett for valuable help and criticism.

REFERENCES

1. Burt, Sir C. (1940). *The Factors of the Mind* (1st ed.). London : University of London Press.
2. Burt, Sir C. (1945). ' The reliability of teachers' assessments of their pupils.' *Brit. J. Educ. Psychol.*, XV, 80–92.
3. Cast, B. M. D. (1939, 1940). ' The efficiency of different methods of marking English composition.' *Brit. J. Educ. Psychol.*, IX, 257–69, X, 49–60.
4. Edgerton, H. A., and Toops, H. A. (1928). ' Formula for finding the average inter-correlation of unranked raw scores without the individual inter-correlations.' *J. Educ. Psychol.*, XIX, 131–8.
5. Finlayson, D. S. (1951). ' The reliability of the marking of essays.' *Brit. J. Educ. Psychol.*, XXI, 126–34.
6. Hoyt, Cyril (1941). ' Test reliability obtained by analysis of variance.' *Psychometrika*, VI, 153–60.
7. Kuder, G. F., and Richardson, M. W. (1937). ' The theory of the estimation of test reliability.' *Psychometrika*, II, 151–60.
8. Snedecor, G. W. (1948). *Statistical Methods* (4th ed.). Iowa : Iowa State Collegiate Press.
9. Tippett, L. H. C. (1941). *The Methods of Statistics* (3rd ed.). London : Williams and Norgate.

Vol. V
Part I

The British Journal of Psychology,
Statistical Section

March,
1952

THE NATURE AND CAUSES OF MALADJUSTMENT AMONG CHILDREN OF SCHOOL AGE

By CYRIL BURT and MARGARET HOWARD
University College, London

I. *Maladjustment as a Psychological Category.* II. *History and Definition of the Term.* III. *The Need for Statistical Verification.* IV. *A Factorial Study of Maladjustment.* V. *The Causes of Maladjustment.* VI. *Summary and Conclusions.*

I. MALADJUSTMENT AS A PSYCHOLOGICAL CATEGORY

The Classification of Handicapped Children. Both psychologists and educationists frequently refer to what they call ' the maladjusted child ' as though such children formed a separate type or class. Nearly all text-books on educational psychology make at least a passing mention of this group. But few indicate how it is related to other groups.

Perhaps the clearest classification is that given by Sandiford in what for long was one of the best-known text-books on educational psychology. In his discussion of individual differences ((14) pp. 178f.) he distinguishes five main groups: (i) ' a homogeneous central group ' consisting of average pupils ; (ii) towards the upper end of the scale, the ' intellectually gifted ' ; (iii) towards the other end (*a*) those who are defective in special intellectual capacities, of whom the most easily recognized are the partially blind, the partially deaf, and those exhibiting defects of speech ; (*b*) the intellectually defective or ' feebleminded ' ; and (*c*) the ' maladjusted.' Among these last three groups the first, as he points out, is plainly composite. As to the second, it had often been alleged that the " mentally defective form as mixed a group as the physically defective " ; but ' correlation studies ' have established the existence of an ' innate factor of general intelligence,' in terms of which the intellectually defective could be described. The ' maladjusted group,' however, are, as he admits, " much less well defined " ; and neither Professor Sandiford, nor, so far as we are aware, any other writer, has succeeded in producing first-hand evidence to support the notion that they form a distinct or relatively homogeneous psychological type.[1]

Reports of education authorities and child guidance clinics often tabulate cases of ' maladjustment ' as distinct from those of ' mental deficiency,' ' educational backwardness,' ' psychoneurosis,' and the like. But their implicit interpretations of the word vary widely. As their figures and cases show, with some writers the ' maladjusted ' consist mainly of ' delinquents ' and ' neurotics,' while with others both these subgroups are placed in separate categories of their own. Most psychiatric

[1] It is almost everywhere recognized, tacitly if not explicitly, that there are no *sharp* lines of demarcation between general and special disabilities, or between intellectual, emotional, physical, and social handicaps, or again between innate intellectual deficiency and acquired intellectual backwardness. Recent writers, who deny the possibility of classifying mental conditions in this way, base their objections mainly on the fact that both cases and causes widely overlap. Undoubtedly they do ; otherwise factor analysis would never have been needed to establish the differences : (no one would carry out a factor analysis of sensory defects to demonstrate that blindness forms a category distinct from deafness). In all such problems, therefore, we can only hope to demonstrate the existence of *relatively* distinct and *relatively* homogeneous psychological groups.

writers who use the term appear to give it a different meaning from that previously adopted by psychologists ; generally they treat it as *including* the ' educationally subnormal ' ((5) pp. 168f.).[1]

The problem has recently acquired considerable practical importance. The Education Act of 1944 enumerates eleven categories of ' handicapped children.' If we leave on one side the various groups of physically defective pupils, and substitute the phrase ' educationally subnormal ' for ' intellectually defective,' the classification corresponds closely with that given by Sandiford and other educationists. For the first time in any statutory enactment ' the maladjusted ' are specified as a separate class. The Act itself does not explain how the term is used. But, according to the Ministry's regulations (12), such cases are to be distinguished from other forms of ' psychological disturbance ' by the fact that they ' require special educational treatment in order to effect their personal, social, or educational readjustment.' But this merely tells us that the ' maladjusted ' child is one who requires to be ' readjusted,' which is not very informative.

The problems to which this enactment gives rise have, not unnaturally, led to numerous discussions. From these it seems clear that there are as yet no agreed conclusions about the probable number of such cases, or the causes responsible for such a condition, or even the precise definition of the term. It is partly on this account that we have been asked to summarize in more accessible form some of the relevant data collected during earlier surveys and hitherto available only in unpublished memoranda, and to amplify in fuller detail some of the conclusions then drawn.[2]

II. HISTORY AND DEFINITION OF THE TERM

History of the Concept. Since the term has been used in so many different senses, no statistical investigation of the subject can have much value unless the investigators first explain how the word is being used. We must therefore begin by formulating a definition of the term, sufficiently precise for practical purposes, and by setting forth, as succinctly as we can, the justification for the interpretation proposed.

Although new in ministerial publications, the notion of maladjustment is by no means new in psychology. The idea of adaptation or adjustment has formed a leading concept with nearly all the representatives of the British evolutionary school.[3] " Mental life," so Spencer repeatedly declares, " consists in a continual *adjustment* of inner processes to outer processes." [4] And Sully adopts much the same language : " Mental development may be viewed as a progressive adjustment of the individual organism to its environment. . . . It is only as such adjustment is effected that the conditions of stable life are realized." He goes on to apply the notion more particularly to define the distinction between the *normal* and the *abnormal.* " There is no wide gap," he says, " between the two, such as the popular contrast suggests. The sound shade off into the unsound by a continuous succession of fine gradations. The normal is adjustment of the organism to its environment ; the abnormal, simply

[1] In this country the most authoritative statement of the views held by leading psychiatrists is that contained in the *Survey of Child Psychology*, edited on behalf of the Child Guidance Council by its Medical Director and consisting exclusively of contributions by well-known medical writers (5). Here, among the four contributors who use the term, the meaning varies appreciably. The concluding chapter (on ' The General Management of Maladjustment in Children ') treats the word as a generic term to cover *all* cases referred to child guidance clinics. Other writers use it to describe specific groups. Dr. Bryan implies that the majority of maladjusted children are dull (" I.Q. 70–85 ") ; Dr. MacCalman states that their mean I.Q. " corresponds closely to that of the average population."

[2] Much of the following article, and the tables accompanying it, were incorporated in a memorandum for a working party appointed by the British Psychological Society to enquire into methods of ascertaining and treating maladjusted children of school age. We gratefully acknowledge permission to use the material here. From the discussions on the memorandum it seemed clear that something like an historical justification for our use of the word ' maladjustment ' would be welcomed, particularly in view of the differences of interpretation between psychiatrists and educational psychologists.

[3] Cf. Murphy, G., *An Historical Introduction to Modern Psychology*, 1949, pp. 109f., and Huxley, J., *Evolution : The Modern Synthesis*, chap. VIII. The word ' adjustment ' has come into use among psychologists to designate relatively permanent adaptations made by the *individual*, partly because the word ' adaptation ' has been adopted by evolutionary writers to designate modifications acquired by the *species* as a result of natural selection, and partly because it has been used by physiological writers to denote the *temporary* adjustments resulting from sensory changes (as in the phrases ' dark adaptation,' ' temperature adaptation,' etc.).

[4] Spencer, H., *Principles of Psychology*, I, 1870, chap. VII, ' The Scope of Psychology ' (pp. 129f.).

maladjustment." Moreover, he insists that " the psychologist's reference to external conditions must above all include that social environment in interaction with which the individual thinks and acts. We cannot understand an individual mind in abstract separation : each thinks, feels, and acts as he does under the constant influence of social relations." [1]

In his *Teachers' Handbook of Psychology* (1886) he makes this doctrine the basis of his account of the mental development of the individual child. In its essential nature, we are told, it consists of " successive adaptive modifications, issuing in a growing adjustment of the individual to the circumstances of his environment, and in a wider range of adaptations." In a modern civilized society, he explains, we may discern three successive stages of increasing range : the adjustment of the child (i) to life in his family, (ii) to life at school, and (iii) to life in the community at large ; while all the time there is an increasing need for (iv) self-adjustment.[2]

Sully's view of mental abnormality was adopted by Mercier, and was even made the basis for his definition of insanity,[3] a term which he used in a broad etymological sense to cover any form of mental ill-health. Such a conception, however, was too far removed from the current medical view of mental illnesses, which considered them to be classifiable into distinctive types according to the causes and the symptoms found *within the patient* [4] ; and Mercier was in consequence vigorously criticized by his colleagues for " substituting vague speculation for clinical observation." Quite recently, however, a number of psychiatric writers in this country have tentatively revived the term. Thus the various medical authorities who have contributed to the *Survey of Child Psychiatry* have used it sometimes to describe the causes of delinquent and psychoneurotic behaviour, occasionally to describe the effects of physiological and biochemical disturbances, and (most frequently of all) as a generic term to cover psychoneurosis, delinquency, and even educational backwardness ((5) pp. 168f.).

Owing to the fact that they seemed to necessitate a teleological interpretation, the concepts of adjustment and maladjustment at first found little favour with American psychologists. But, since the decline of the more extreme behaviourist doctrines, several writers have sought to reintroduce a biological standpoint, and have thus made increasing use of these terms.[5] Allport, for example, in setting forth what he terms ' the biological theory of personality ' makes ' adjustment ' a key word.[6] And a similar change in standpoint has appeared in American psychiatry.[7]

[1] Sully, J., *The Human Mind : a Textbook of Psychology*, 1890, I, pp. 34, 203.
[2] *Op. cit.*, 4th ed., pp. 66, 529, 572f. It may be noted that the *Survey of Child Psychiatry* adopts the same threefold stages in its successive chapters on ' Maladjustment ' ((5) pp. 153–86).
[3] Insanity consists in " the non-adjustment of the mental relation to the relation in the environment." The ' ability to adjust ' he designates ' intelligence ' ; and hence his final definition of insanity is ' a disorder of intelligence ' (*The Nervous System and the Mind*, 1888). As was immediately pointed out by psychological critics (*Mind*, XIII, pp. 263f.), Mercier's materialistic basis, together with his rejection of introspective terms, led him to give an excessively intellectualistic bias to his theories, and to ignore the affective and conative aspects of mental disorders, which had been so strongly emphasized by Sully.
[4] See, for example, the ' suggested classifications ' described by Henderson and Gillespie in their text-book ((9) pp. 20f.) ; their own ' scheme for the examination of the patient ' makes no reference to the environment (pp. 67f.) ; the two pages on ' environmental influences ' in the chapter on ' psychopathology ' merely summarize the Freudian account of the growth of the ego-ideal. In their chapter on ' The Psychiatry of Childhood ' (pp. 488–536) they classify cases needing psychiatric study into nine main ' categories ' ; but none of these corresponds with that of the ' maladjusted child.' Barton Hall refers in passing to the inclusion of the maladjusted child in the Education Act of 1944 ((6) pp. 8 and 335), and has a section on ' environmental adjustment ' as consisting in ' common-sense measures ' for ' rectifying circumstances in the child's home, school, or neighbourhood,' but does not treat of the maladjusted child as a distinguishable type.
[5] The change is strikingly illustrated in the successive versions of the admirable text-book by Boring, Langfeld, and Weld. The chapter on ' Personality ' in *Psychology : A Factual Textbook* (1935) briefly and critically referred to ' the difficult concept of adjustment ' (chiefly in connexion with the definition of intelligence) and finally dismissed it as unhelpful. In *Foundations of Psychology* (1948) there is now a whole chapter of 35 pages devoted explicitly to ' Personal Adjustment.'
[6] ((1) pp. 114f.) The word also appears in his own definition of ' personality ' (p. 48). However, later in the book he distinguishes and contrasts the ' adaptive ' and the ' expressive behaviour ' of the personality (p. 465).
[7] Cf. Shryock, R. H., 'A Delayed Advance Against Mental Disease' ((15) pp. 305f. and refs.) and Meyer, A., ' British Influences in Psychiatry and Mental Hygiene : 14th Maudsley Lecture,' *J. Ment. Sci.*, LXXIX, 1933, pp. 435–63.

Not only British writers,[1] but also American, have frequently complained that, until quite recently, psychiatry in that country so insistently ignored the influence of social and psychological factors in the causation of mental disorders ; and, " largely because child guidance in America was, during its earlier stages, developed by psychiatrists rather than by psychologists or educationists, the importance of the sociological aspects remained " (so we are told) " for long unrealized." Reymert, for instance, points out that, at the time of Healy's early studies, " the clinic staff did not include a social worker " : " the work [on mental subnormality] had expanded gradually from a consideration of the mentally retarded to that of the educationally retarded " ; and only in the later development of the clinical study of children was it recognized that " the feeble-minded and the maladjusted are separate problems," and that " educational maladjustment is often only a part of a picture of general maladjustment." In child psychiatry, as in other branches, this change of viewpoint is due in great measure to the teaching of Adolf Meyer and the ' school of psychobiology, which insisted that the psychiatrist is concerned " with the total individual in relation to the total situation."[2]

External and Internal Maladjustment. Unfortunately, both in America and in this country, those psychologists who use words like ' adaptation,' ' adjustment,' and ' maladjustment ' seldom offer any explicit definitions of the meanings they attach to them ; and psychiatric writers usually appear to employ them in a popular rather than a technical sense. Often, indeed, they evidently assume that the terms refer to what may be called ' internal ' as distinguished from ' external maladjustment.' [3]

Thus teachers are not infrequently told that " the maladjusted child is the child whose personality is ill-organized or badly integrated : he is a machine whose parts do not fit ; and, as with other machines in that condition, he must be sent to a repair-shop in order that the appropriate expert may effect the necessary readjustments." From the very nature of their subjects, it is said, " psychologists and physiologists are primarily occupied with the study of distinct functions in isolation, whereas psychiatrists are concerned with the study of the internal organization or ' integration ' of the personality as a whole." Such criticisms are quite justifiable if (as seems to be the case) the critics are referring chiefly to the attitude of Titchener and his American followers, who persistently deprecated all attempts to apply psychology to social or practical problems. But they are scarcely true of British

[1] McDougall, writing in 1935, refers to American psychiatry as " a combination of Watson and Freud " (*Psychoanalysis and Social Psychology*, p. 12) ; Murphy, writing of " the clinical psychology of the second quarter of this century," makes almost exactly the same remark (*An Historical Introduction to Modern Psychology*, p. 419) ; cf. also Zilboorg, G., *A History of Medical Psychology* (1941).

[2] Harms, E. (ed.), *Handbook of Child Guidance*, 1947, pp. 225–7. Reymert, however, points out that " both William James and Stanley Hall had represented the point of view that behaviour is after all an expression of the adjustment of the individual to the environment." The reader of the *Handbook* is perhaps a little astonished to find that only one chapter out of 41 deals explicitly with ' maladjusted children,' and there the phrase is used merely as a euphemism for ' delinquents.'

In American psychology the revival of the concept appears to have been due to three main influences : first, the development of purposive behaviourism, which employs adjustment (as the early trial and error school had done) as a useful concept in theories of learning ; secondly, the popularity of the field theories of Lewin, especially as applied to child behaviour ; and thirdly, the growing interest in the cultural aspects of anthropology, as developed in this country by McDougall and Malinowski and more recently in America by Ruth Benedict and Margaret Mead : cf. also (11).

[3] Any definition of ' maladjustment ' implies a prior definition of ' adjustment.' Most dictionaries mention two relevant senses : (1) ' a thing is adjusted when as a whole it conforms with *external* conditions ' ; (2) ' a thing is adjusted when its parts conform *with one another*.' The word itself is derived from the Latin *ad-juxtare*. Hence etymologically the former is the original and the more correct interpretation. The second meaning is apparently due to the idea that the word is related to *justus* (' fit, right, or proper ') : but *adjustare* would be an impossible formulation. Certainly, in biological and psychological writings, the term, as originally introduced, bore the first of these two meanings, though no doubt appropriate adjustments to an external situation cannot be carried out when the necessary internal co-ordinations become disordered. In any case, it is the maladjusted *behaviour* which we actually *observe*, and to which reference must be made in any operational definition ; such behaviour is primarily a matter of *external* adjustment. The further assumption of various types of internal maladjustment (' disintegration,' ' conflict,' ' frustration,' ' loss of equilibrium,' etc.) are merely hypotheses introduced by various theoretical schools—hypotheses which may or may not be helpful in enabling us to understand this or that particular case.

psychology.[1] Ward, Sully, Stout, McDougall, and their followers have never tired of emphasizing "the unity of the personality." And they go further still. Again and again they remind us that the personality itself is an abstraction, and can never be adequately studied apart from the psychological environment. Indeed, it was to emphasize this latter point that terms like 'adaptation' and 'adjustment' were employed.

The Concept of Maladjustment in Child Psychology. Here we shall be concerned almost exclusively with 'maladjustment' as a term of child psychology. In this country the foundations of child study and child guidance were laid by Sully; and consequently his conception of adjustment and maladjustment has figured prominently in British writings on the subject from the very start.

As Sully himself repeatedly affirmed, the term was meant to stress the fact that the maladjusted child is not to be considered as a patient seen in isolation at the clinic consulting room: he is to be regarded as an individual personality interacting with the immediate bio-sociological environment. To borrow the language of 'dynamic psychology' which McDougall made popular, the 'mental forces in the field' must be taken into consideration quite as much as the 'mental forces within the individual.' It follows that, in any statistical investigation on the problem of the so-called maladjusted child, it will be essential to secure data, not only in regard to the child's personal characteristics—his general and special abilities, his health and physique, and his temperamental, emotional, and moral characteristics, but also in regard to his past and present environment, at home, at school, at work (if he is already employed), and during his hours of play or recreation.

Definition. As we have seen, the appearance of this new category in official regulations has led to an increased number of speculative statements regarding the nature and causes of the conditions falling within its scope. Very little, however, seems so far available in the shape of empirically verified facts. Probably an acceptable definition of what the term implies can only be reached at the end of a systematic enquiry into the whole problem. But some preliminary formulation is necessary at the outset. For educational purposes, therefore, we provisionally suggest that a maladjusted child may be defined as one *whose adjustments to the recurrent situations of his everyday life are less adequate than might reasonably be expected from a child of his mental age, and whose condition or circumstances therefore require special study and treatment.*

The use of the term 'recurrent' is intended to imply that the maladjustment is not a temporary consequence of some passing crisis, but prolonged and relatively permanent. The recurrent situations will involve, on the one hand, (1*a*) the physical and perceptible needs of the child's material environment and (1*b*) the social and moral needs of his human environment, and, on the other, (2*a*) the conscious or non-conscious needs of the child himself and (2*b*) his capacity for fulfilling those needs. His adjustment may be described as 'inadequate' when these needs, as a whole, are not effectively fulfilled.[2]

[1] The criticisms we have quoted are frequently urged by those British psychiatrists who have been largely influenced by American views and seem more familiar with American psychology than British. Thus a common complaint is that the academic psychologist is concerned chiefly with "the endless study of isolated performances of sensory, motor, and other functions in the laboratory"; whereas the practical psychiatrist is forced to "treat the whole patient": he cannot, we are told, "escape the necessity of a knowledge of the structure and function of the tangible mechanism as a whole, and this, of course, entails a medical training" (cf. (8) pp. 14, 23, 177f.). Unfortunately these writers seldom specify any psychological books or other publications by name. But it seems clear that they have chiefly in mind what Murphy calls "the structuralist school," which, under the acknowledged leadership of Titchener, dominated American work down to about 1910.

[2] All these phrases need a fuller explanation than is possible here. To indicate the view they are meant to imply, we may refer to the discussion of 'adjustment' in Stout's chapter on conative activity (*Analytic Psychology*, I, pp. 146f.). Those who dislike the teleological implication of the phrase 'fulfilling needs' may adopt Stout's conception of 'the reattainment of relatively stable equilibrium' (p. 151), although we should agree with him that in mental dynamics 'equilibrium' means something more, or at any rate something more specialized, than the equilibrium of purely material systems.

At its simplest the postulates required by such a definition are those suggested by early experiments on ' adaptive behaviour,' like those of Lloyd Morgan and Hobhouse. We assume (i) that the individual possesses certain conative tendencies, inborn or acquired, which (under certain internal conditions) are aroused by some specific change in the external environment, and cease when some specific end-result is reached, i.e., when some need is satisfied ; (ii) that the external situation not only arouses the conative impulse, but also impedes its immediate fulfilment by any automatic reaction, such as might result from pre-existing instinctive or habitual patterns of behaviour ; (iii) that the immediate result of the external stimulation and obstruction is a state of disequilibrium or tension : (in metaphorical language we may say that the flow of energy is at once increased and temporarily blocked) ; (iv) that, in such a situation, the internal structure of the individual organism enables it to develop *alternative* patterns of activity (based on ' trial-and-error ' or ' discriminative insight ') ; and (v) that, after a period of varied activity, and as a result of repetitive trial, discriminative insight, and eventual success (or some equivalent process), a *new* mode of reaction is eventually acquired, which in some degree provides a more or less adequate solution to the problem.[1] It follows that, if the psychologist is seeking to aid a child in a condition of ' maladjustment,' the assistance should take the form rather of ' training ' the child to meet the situation (e.g., by assisting discrimination, encouraging alternative reactions, or simplifying the situation itself) than of medical ' treatment.' Adequacy is plainly a matter of degree ; hence the need for the second clause in the definition above. Two main grades may conveniently be distinguished : (i) if the child merely requires individual assistance from the teacher, social worker, family doctor, or other adviser, we suggest that the condition may be described as one of *mild* maladjustment ; (ii) if he requires the attention of a specially qualified expert (e.g., a psychologist at a child guidance centre), it may be described as *severe*. In a large mixed urban area, like London, surveys suggest that roughly 10 per cent. of the school population fall into the first grade, and 2 per cent. into the latter.[2]

III. THE NEED FOR STATISTICAL VERIFICATION

Alternative Methods of Proof. Most of the descriptions given of the various conditions which are assumed to lead to maladjustment appear to be based on more or less plausible deductions from the writers' psychological preconceptions. It is our view that such assumptions are of little scientific value until they have been critically checked and verified by some sort of objective test. However, most writers on child psychiatry apparently hold that any attempt at statistical confirmation is not merely invalid but actually impossible ; in their opinion the only satisfactory mode of verification is the first-hand intuitive insight of the experienced practitioner actually handling such cases.

This point of view has been forcibly argued by a former Medical Director of the Child Guidance Council, and is frequently expressed by other writers on the subject.[3] They quote with approval the

[1] We have here for the most part adopted the terminology used by McDougall in his description of what he prefers to call ' adaptive achievement ' (e.g., in *The Energies of Men*, 1932, chap. XXIII). Such concepts as ' conative tendencies ' (or ' motives '), ' ends,' ' satisfaction,' ' discriminative insight,' etc., were severely criticized by American behaviourists ; but have reappeared in more recent writings. Thus, in the *Textbook* already cited by Boring, Langfeld, and Weld, adjustment is defined as " a persistent activity directed towards the realization of goals in the face of difficulties " or (a little more generally) as " the process by which a living organism maintains a balance between its needs and the circumstances influencing their satisfaction." And, it is now freely admitted, " all action is motivated and purposive " (pp. 511, 514).

[2] This figure is based on the assumption that not only ' educationally subnormal children ' (the dull, backward, or mentally defective), but also the delinquent, will be enumerated in separate categories of their own. If, however (as the Ministry's regulations would seem to imply), the delinquent are to be included among the maladjusted, then the proportion must be increased to at least 3 per cent. The difference between ' mild ' and ' severe ' degrees of maladjustment correspond to the differences between ' mild ' and ' severe ' retardation in intelligence, as connoted by the phrases ' dull ' and ' mentally defective ' respectively, and could indeed be defined in terms of a mental age for adjustment. The two groups together would seem to be roughly covered by Carl Rogers' category of ' seriously maladjusted,' in which the maladjustment, though ' serious,' is not necessarily ' severe.' Rogers' criterion, however, was somewhat different, namely, maladjustment in at least four different directions as specified by his various criteria, which to some extent depended on general and special intellectual disabilities (13).

[3] Maberly, A., ' The Personality of the Problem Child,' *Brit. J. Educ. Psychol.*, XVI, 1946, pp. 103f. ; see also the correspondence columns of the *Brit. Med. Journ.*

conclusion of one of the contributors to the *Survey of Child Psychiatry*. " Any attempt to deal with the matter statistically is entirely unconvincing, so that in these matters the guesses of a clinician must be allowed more value than the statistician's argument." [1]

We fully agree that " statistics alone can prove nothing," and that in the past statistical investigators have furnished some justification for such criticisms by confining themselves to " tables of figures and brief supplementary comments." Indeed, it is for that reason that we have prefaced our statistical analysis by a discussion of current views. But before we turn to the results of our own enquiries, our critics may feel that we should offer some brief defence of the procedure we propose. Perhaps the best way of meeting such objections is to examine the alternative method of proof, which is customarily followed. It relies rather on plausible generalization than on systematic verification.

In the opening chapter of the *Survey* just quoted, an attempt is made to demonstrate the importance of " the physiological approach to psychiatric problems," which earlier ' child psychologists ' were held to have neglected. To prove his case, the author [2] cites the history of a girl of 10 referred for persistently stealing sweets or money for sweets. Being ' thin and lackadaisical,' her blood-sugar was examined. Found to suffer from hypoglycæmia, she was put on a diet including regular supplies of glucose, and " from that moment never stole anything again." Now, " when the blood-sugar level sinks, profound changes in cortical function may ensue, amounting to coma and death." Accordingly (we are told) " *it is clear* that the low blood-sugar created a tension in the mesencephalic and diencephalic centres, *which was transmitted to the effector areas* in the frontal lobes." Hence the thefts. At this point, however, a statistical argument (a little disguised) persists in breaking in. Numerous patients have been found suffering from hypoglycæmia : do all of them (the reader naturally enquires) relieve this ' mesencephalic tension ' by stealing, or, if not, how many do ? The writer recognizes this gap in the proof ; and adds " we must postulate *some specially directed* spread of excitation here, for every case of hypoglycæmia does not steal sweets. Nevertheless, the *chief* stimulus was apparently the hypoglycæmia " ((5) pp. 12f., our italics).

But, of course, there are a good many alternative hypotheses, far more simple and far less speculative, which would equally well explain the ' cure,' and which could perhaps have been partly confirmed, had some information been included about the general background of the case. Moreover, since the writer's object was to demonstrate the importance of ' biochemical needs ' (as distinct from ' biological needs ') in precipitating disorders of conduct, it would surely be helpful to learn in how many other cases, where habitual pilfering has been prompted by a childish desire for sweets, a similar hypoglycæmic condition has been discovered.

We do not for one moment wish to deny the possibility that biochemical conditions may profoundly influence the child's mental state. We merely contend that, before a single case of this kind can be regarded as typical and the proposed explanation accepted as proved, a statistical investigation is essential : we have first to show that the causes found are accompanied by the effects alleged with a frequency which cannot be explained by coincidence, and secondly to determine how important or how common such a factor may be.

Sources of Data. In London it was part of the duties of the Council's psychologist to examine pupils whose mental development appeared to be hampered in any way, whether by personal or by environmental difficulties. To determine the number and types of children needing special investigation or treatment, surveys were carried out from time to time in different areas ; and instructions were issued to teachers, care committee workers, and school medical officers, explaining what kinds of pupils should be referred. In addition to other types of case, it was suggested that they might submit for special examination " children whose school progress, behaviour, or emotional condition appeared to indicate that they might be handicapped by some form of maladjustment." The term was briefly explained along the lines indicated above ; and special emphasis was laid on the fact that the maladjustment might be due, not only to conditions in the child himself, but also to those of his home, school, or everyday life (cf. (5) p. 15).

[1] ((5) p. 31.) Dr. Winnicott's contribution dealt with ' the psychology of juvenile rheumatism,' and it seems a little doubtful whether he meant his conclusion to be extended to *all* problems of child psychiatry, as is commonly supposed.

[2] The author, who also edits the *Survey*, was medical director of the Child Guidance Council, and therefore speaks with authority as well as from wide experience. We select this volume because it is one of the few addressed to professional rather than to popular readers ; and is based on the Council's desire to " collect the knowledge of British experts and cover the general field of child psychiatry." We may note that, in the only chapter which presents any detailed statistical evidence (that dealing with ' Psychoses in Children ' by Dr. Gillespie), the figures are derived almost entirely from continental investigators.

There were practical reasons for the inclusion of some such category. The earlier surveys had dealt chiefly with those who were actually backward, delinquent, defective, or otherwise subnormal. But terms like ' defective ' or ' subnormal ' had come to imply some pathological characteristic in the child—some intrinsic weakness or infirmity of the mind, rather than the effects of the constant interaction between the child on the one hand and his present or past circumstances on the other. In the absence of any such mental defect or illness, teachers were apt to assume that every pupil ought to adjust himself to the work of his class ; and many failed to realize that often the only effective procedure was to adjust the choice of class or teaching methods to the needs of the child. Further, it became clear that the work of the psychologist should be preventive as well as curative. It therefore seemed urgently desirable to investigate, at the earliest possible stage, cases of *potential* backwardness, delinquency, or neurosis. However, to ask teachers and others to notify not only actual but also potential delinquents and neurotics would have been useless. The teacher regarded the discovery of neurotic conditions a matter for the doctor ; the doctor considered his task to be the diagnosis of actual cases ; and to pick out certain children as ' potential delinquents ' seemed to be condemning them before they had committed any offence. Above all, it was desired to avoid the common misunderstanding that a child referred to the Council's psychologist (or ' mental expert,' as he was sometimes described) must necessarily be either ' mentally defective ' (in the technical sense of defective) or ' mentally ill ' (in the strict sense of illness). ' Maladjustment '—the term so often used by Sully and other British educationists, and already part of every teacher's vocabulary—appeared to cover most of these points.[1]

In addition to investigating cases referred by teachers and others, a careful watch was kept for similar cases during periodical surveys of the school population ; and the data that we propose to analyse here are drawn mainly from these two sources.[2] Since it appeared conceivable that the term ' maladjustment ' might be interpreted differently by other investigators, we have also endeavoured to secure similar information about children who have been more recently referred to child guidance centres, and have then come under our own notice or that of other co-workers later on, usually in the course of researches upon some other topic.[3] No significant differences seem discernible between the data reported for these various groups of cases. On the whole, therefore, we are inclined to conclude that, notwithstanding differences in theoretical standpoint, in actual practice the term is now used, at any rate in London, in much the same way.

It should be added that children referred solely or primarily on the ground of actual educational backwardness or actual delinquency will not be included in the present series. A number of the cases still retained may no doubt have been sent forward partly because of undisciplined behaviour either at home or at school ; others proved on investigation to have committed offences already, generally of a minor type ; and a small proportion turned out to be educationally subnormal as well as maladjusted. But in every case considered in the following analysis, the

[1] Among American writers on child guidance the phrase ' problem child ' seems to be used for much the same reasons. The *Handbook of Child Guidance*, for example, classifies cases needing guidance as (i) normal, (ii) subnormal, (iii) ' problem children,' and (iv) the physically handicapped (a scheme very similar to Sandiford's). But surely we need a more specific term which will indicate, at least provisionally, the *nature* of the ' problem ' envisaged ; and further the phrase still suggests that the ' problem ' is a characteristic solely of the child.

As we shall discover later on, the term maladjustment is itself not very suitable for any *final* psychological classification. There is, too, the common difficulty that attends all such classification : are we trying to classify by symptoms or by causes ? Maladjustment, as we have seen, may refer to either.

[2] The analysis of case-histories and the greater part of the calculations have been carried out by Miss Howard. We have to express our thanks to the various teachers, care committee workers, school medical officers, and psychologists who assisted in the original investigation of the individual cases. Some of the data used in the following paper have been taken from an unpublished L.C.C. Report on *Subnormal Children of School Age* and other surveys. Professor Burt therefore wishes to acknowledge his indebtedness to the Education Committee for permission to make use of material collected while he was working in the Education Department.

[3] In a further endeavour to compare our own cases with those studied by other investigators, we have examined numerous case-histories reported in the various journals of medical psychology. We have been astonished to discover how rarely any precise or detailed information was given in regard to environmental conditions. Even when mentioned, the descriptions are so brief and vague that, more often than not, the data are quite useless for any systematic comparison.

TABLE I. CONDITIONS ASSOCIATED WITH MALADJUSTMENT
GROUP FACTOR-SATURATIONS

FACTOR	I	II	III	IV	V	VI
I A. *School*						
Absence	·19	·53	·17	·09	·14	·05
Inappropriate School	·26	·42	·10	·06	·08	—·07
Inappropriate Class	·32	·68	·08	·04	·22	·03
Uncongenial Teacher	·38	·71	·14	·13	·17	·06
Average	·29	·58	·12	·08	·15	·02
I B. *Home*						
Economic Condition	·27	·28	·38	·07	·09	·02
Overcrowding	·19	·16	·46	·03	·07	·04
Parent-Child Relation	·31	·11	·75	·12	·04	·19
Parental Relations	·43	·09	·59	—·09	—·07	·08
Defective Relations	·26	·07	·53	·08	·05	·17
Defective Discipline	·35	·18	·68	·04	—·08	·07
Average	·30	·15	·57	·04	·02	·09
I C. *Out of School and Home*						
Conditions at Work	·17	·22	·29	—·07	·06	—·07
Conditions during Leisure	·31	·24	·41	·06	·08	—·08
Average	·24	·23	·35	—·01	·07	—·08
II A. *Physical*						
Developmental	·18	·06	·09	·28	·06	·18
Pathological	·11	·07	·03	·21	·12	·13
Average	·15	·06	·06	·24	·09	·16
II B. *Intellectual*						
General Intelligence	·21	·13	·11	·26	·61	·03
Special Disabilities	·27	·18	·07	·31	·41	·07
Inferior Attainments	·44	·37	·12	·42	·53	·05
Average	·31	·23	·10	·33	·52	·05
II C. *Emotional*						
General Emotionality	·41	·16	·09	·37	·13	·64
Extraversion	·26	·13	·08	·29	·11	·13
Introversion	·37	·07	·13	·17	·06	·57
Adolescent Instability	·18	·12	·07	·26	·03	·43
Alleged Neurosis	·22	·11	·12	·22	·07	·41
Alleged Delinquency	·16	·07	·18	·34	·11	·25
Conflict, etc.	·39	·05	·10	·29	·09	·32
Average	·28	·10	·11	·28	·08	·39

ostensible ground of reference was some form of maladjustment as interpreted above.

From what we have said, it will be obvious that the group with which we are concerned is somewhat vaguely defined. It will be our task to analyse its nature more precisely, to discover what causes are mainly responsible for its existence, and if necessary to suggest some more exact method of classification.

IV. A FACTORIAL STUDY OF MALADJUSTMENT

The Chief Characteristics of the Maladjusted Child. The first problem that calls for investigation is the validity of classifying certain children in this way. To what extent do those reported as suffering from maladjustment constitute a homogeneous

class, or form a distinct psychological category of their own? Is there any definite attribute or component, no matter how complex, which underlies all these cases, and is characteristic of no other group? To answer such questions the most obvious procedure is to undertake a factor analysis for the main conditions reported. Since, as we have seen, the term necessarily refers to the relations or interactions between two sets of conditions—environmental on the one hand and personal on the other—it is necessary to collect detailed information as regards (I) the material and psychological conditions in the child's home, school, and places of business or recreation, and (II) the child's own health, intellectual, emotional, and moral characteristics, especially in so far as they affect his reactions to his environmental conditions.

Data. Fairly complete information was available for 273 cases. These were mainly children between the ages of 9·0 and 13·0. No significant differences were found between boys and girls; and the groups were too small to permit of subdivision according to age. The conditions ascertained were classified under 24 subheadings, as shown in Table I. For most of the personal characteristics (intellectual abilities and attainments measured by tests and emotional characteristics by a rating scale on the basis of direct observations), and for several of the environmental characteristics, the data obtained consisted of graded assessments; but for the majority of the latter the information related merely to the presence or absence of some specified condition. Accordingly, the inter-correlations were calculated by the biserial, tetrachoric, or full product-moment method, according to the nature of the variates paired. The entire 24 × 24 table of coefficients was then subjected to analysis first by simple summation and then by the group factor method, the final set of group factors being obtained by an arithmetical rotation: (for details of the procedure used see this *Journal*, III, pp. 50, 75). The saturations thus eventually reached are set out in Table I.

To reproduce the entire 24 × 24 correlation table as it stands would take up an undue amount of space, and would not be very informative. We therefore print only a condensed correlation table, containing the average coefficients for the six main groups of conditions (Table II) and the average factor-saturations obtained by the bipolar and group factor procedures from the complete correlation table (Table III). The sixfold subdivision of the 24 conditions was based on the detailed results shown in Table I. From the average coefficients given in Table II it will be seen that the correlations

TABLE II. AVERAGE CORRELATIONS

CONDITIONS	I A	I B	I C	II A	II B	II C
I A. *School*	(·470)	·251	·255	·123	·339	·197
I B. *Home*	·251	(·443)	·298	·114	·212	·208
I C. *Out of School*	·255	·298	(·244)	·065	·192	·105
II A. *Physical*	·123	·114	·065	(·121)	·203	·189
II B. *Intellectual*	·339	·212	·192	·203	(·534)	·278
II C. *Emotional*	·197	·208	·105	·189	·278	(·342)

TABLE III. AVERAGE FACTOR-SATURATIONS

	BIPOLAR				GROUP					
FACTOR	I	II	III	IV	I	II	III	IV	V	VI
I A. *School*	·547	·161	·322	·203	·287	·585	·123	·080	·152	·018
I B. *Home*	·551	·283	−·225	−·113	·302	·148	·566	·042	·017	·095
I C. *Out of School*	·394	·290	·029	−·073	·240	·231	·352	−·005	·070	−·076
II A. *Physical*	·291	−·177	−·030	−·011	·146	·064	·060	·245	·092	·156
II B. *Intellectual*	·579	−·214	·321	−·222	·307	·227	·098	·330	·517	·049
II C. *Emotional*	·490	−·274	−·132	·098	·284	·097	·111	·277	·085	·392

throughout were rather low, though nearly all positive. With 273 cases any correlation over 0·038 (or rather more where the tetrachoric procedure has been used) may be regarded as statistically significant.

Results. Four factors were extracted by simple summation. These contribute 25, 7, 5, and 3 per cent. respectively to the variance. Thus the first or general factor accounts for far less of the total variance than is ordinarily the case. The second factor divides the conditions assessed into two broad groups, and suggests the existence of two fairly distinct types of maladjustment—(I) those due to *environmental* deficiencies and (II) those due to *personal* deficiencies. The third factor divides the former group into (I A) cases where the deficiencies arise mainly in the *school* and (I B) those where they arise mainly in the *home*, and the second group into (II A) those where the deficiencies are mainly *intellectual* and (II B) those where they are mainly *emotional*. It will be noted that, so far as this third factor is concerned, deficiencies in school adjustment tend to be associated with intellectual disabilities, and deficiencies in home adjustment with emotional disabilities. The last factor, however, tends in some degree to reverse these two associations—i.e., it implies the existence of a small number of cases where the maladjustment due to school conditions is *not* associated with intellectual disabilities, and where maladjustment due to home conditions is *not* associated with emotional disabilities.

On referring to our record cards we found that about 60 per cent. of the cases could be sorted out into classes and sub-classes based on this classificatory system. The remainder show a good deal of overlapping. But, of course, it is precisely when the lines of classification are blurred and obscure that a factorial analysis is needed to establish the most appropriate scheme.

The detailed factor-saturations obtained by the group factor procedure (set out in Table I, and averaged in the last six columns of Table III) are perhaps more instructive, and deserve a fuller study.

(1) There is first a small *basic factor* common to all the cases ; but its largest factor-saturation is far smaller than that found in the various group factors. In a corresponding analysis undertaken for the group of normal cases, much the same factor appears, although with them it contributes even less to the total variance.[1] This factor, therefore, would seem merely to indicate the familiar fact that, both in the general population and in restricted samples from it, there is a small tendency for good conditions to go together and for poor conditions to go together, no matter whether those conditions are environmental or personal, emotional or intellectual. Thus the presence of this basic factor does not indicate the existence of a distinct and sharply demarcated category of maladjusted children : it merely implies that, so far as there is a general factor making for maladjustment, the characteristics of those who are definitely maladjusted differ from the characteristics of those who are normal solely in degree. What really distinguishes the former set of cases, therefore, are certain specific features, which vary from one subgroup to another, rather than any single common characteristics. However, if, for practical or administrative reasons, it is found convenient to place all cases of maladjustment in a separate category of their own, there can be no objection to such a classification on psychological grounds. But the term plainly designates no uniform psychological category or type ; and to ask for a clear-cut picture of ' *the* maladjusted child ' as such is likely to lead to misconceptions.

(2) The first of the narrower group factors appears to be a factor relating primarily to maladjustment at *school*. The highest saturations are for ' inappropriate school class ' and ' uncongenial class teacher ' (0·68 and 0·71). The factor shows some degree of overlapping with certain other characteristics ; among these the largest saturation is that found for inferior school attainments (0·37).

(3) The second group factor, still more clearly marked, refers to maladjustment in the *home*. Here the largest saturations are found for the relations between the child and his parents and for disciplinary conditions in the home (0·75 and 0·68). A moderately high saturation (0·41) will be noted for conditions obtaining out of the home during the child's leisure hours.

[1] (4) p. 5, Table I, ' Correlations between Assessments for Environmental, Physical, and Mental Conditions ' : as there noted, both with normal and subnormal groups, " all good features tend to go together ; this holds true of the features of the environment as well as those of the personality." The same phenomenon appears if, instead of ' correlating traits,' we ' correlate persons ' (or groups of persons) : were there a sharply demarcated category of maladjusted children, the distinction would be revealed by a clear-cut bipolar factor, separating the well adjusted from the maladjusted. But, even when present, such a factor nearly always contributes very little to the total variance as compared with the general factor.

(4) A broader group factor with rather low saturations (col. IV) appears to mark off cases in which the maladjustment springs chiefly from the child's *personal* characteristics. Here the highest saturations are for inferior school attainments (0·42) and general emotionality (0·37). This factor is subdivided by two very narrow group factors.

(5) Of these factors, the first evidently groups together those cases with inadequate *intellectua* capacities, notably retarded general intelligence (saturation 0·61).

(6) The second groups together those cases characterized chiefly by their *emotional* traits, especially high general emotionality and introverted tendencies (saturations 0·64 and 0·57).

There is no discernible group factor for defective physical characteristics, either developmental or pathological.

Analysis of Symptoms. Since many writers discuss ' symptoms of maladjustment ' as though they formed a distinctive syndrome marked by emotional and neurotic disturbances, we have also attempted a factorial study of the symptoms falling under these heads. The results again show that there is no *single* syndrome, distinguishing the group as a whole. On the contrary, so far as the observable behaviour of the child is concerned, maladjustment may show itself in a number of alternative ways. The factors obtained suggest the following reaction types : (i) *withdrawal*, i.e., attempts to avoid the situations chiefly responsible for the maladjustment ; (ii) *repression* of impulses or desires which cannot readily be adjusted to the child's situation ; (iii) *persistence or reinforcement* of inadequate reactions (e.g., failure to change habitual reactions when the external situation changes) ; (iv) attempts at *readjustment at a lower or alternative level ;* these are mainly inferior adjustments which tend to satisfy the requirements of the person rather than of the situation (e.g., a search for compensatory substitutes either in reality or in fantasy) ; (v) excessive *dependence* on some other personality (mother, friend, teacher, or even a casual acquaintance) ; (vi) a number of *miscellaneous reactions* of a primitive type (stammering, incontinence, irritable temper, nightmares, physical symptoms of anger or anxiety, etc.).[1]

The conclusions to be drawn from the foregoing analyses seem clear. The cases put forward as cases of maladjustment form a decidedly heterogeneous collection. Children who are maladjusted in one respect are not necessarily maladjusted in others ; and, no matter what the primary cause may be, the fact of maladjustment, so far as it is revealed by the child's behaviour, may manifest itself by a number of alternative reactions, differing from case to case. The requisite treatment must take this into account. Those who are ' maladjusted ' because the work at school is too hard require entirely different measures from those who are unhappy at home or those whose difficulties are emotional rather than intellectual ; those who respond actively by inferior or undesirable adjustments require a different type of training or treatment from those who react by withdrawal or passive regression.

V. THE CAUSES OF MALADJUSTMENT

Data. As in investigating the causes of juvenile delinquency or educational backwardness, our main procedure has been to compare the conditions observed in cases of maladjustment with those observed in a parallel control group of well-adjusted pupils. To a large extent the control groups used in the earlier investigations were employed for the present study. The figures in the tables which follow are based on data obtained for 394 maladjusted boys and girls, aged 6 to 14, and an equal number of normal children, of corresponding age and sex, and drawn for the most part from the same schools.

For these cases information has been systematically collected according to the schedules described in the reports on juvenile delinquents and backward pupils. Accordingly, there will be no need to repeat the accounts there given of the methods used or to explain in further detail the meanings attached to the various headings. These will be found fully discussed in the publications on these subjects (2, 3).

[1] See (4), p. 15, Table II, ' Factor Saturations for Neurotic Symptoms,' which gives figures for 47 commoner symptoms. Such analyses, however, are not very instructive : to a great extent they depend on the particular symptoms which attract the notice of the teachers, doctors, or social workers who refer the cases for investigation, and very largely reflect the fashionable theories of the moment.

Results. In Table IV we have brought together figures showing the frequency of the more relevant conditions in both the maladjusted and the normal groups. However, the reader may not find it easy to judge the relative importance of each causal condition when the results are expressed in the form of two proportions. Hence, to obtain a single index-figure, we have calculated a point-distribution correlation based on the assumption that the samples to be compared are equal in number (see appendix, p. 58). Any coefficient numerically exceeding 0·10 may be taken as statistically significant. The more conspicuous items under each of the various headings call for a word or two of comment.

One point should be emphasized at the outset. Merely to establish a significant correlation between maladjustment and some particular condition does not of itself prove that this condition has operated as a cause of maladjustment. On the other hand, the fact that the correlation falls so near to zero as to be non-significant, even with large samples like the present, would yield strong presumptive evidence that the condition is not an important cause. As we have already seen, in discussions on maladjustment the favourite method of demonstrating causation is to deduce or at least to illustrate, from a few first-hand case-studies, some intelligible process by which the alleged cause is likely to have resulted in maladjustment. It will now, we think, be clear that such arguments can seldom be convincing, particularly when (as is so often asserted) the process postulated is assumed to be unconscious. What is needed is a combination of the case-study method with the statistical method, such as one of us has already attempted in the case of mental deficiency, educational disabilities, and juvenile delinquency ((2), (3), and refs.). But perhaps the most conclusive proof is that afforded by the eventual failure or success of specific methods of treatment, directed to remove or counteract this or that presumable cause. However, a completely cogent demonstration can rarely be secured except in a carefully planned investigation, where the conditions can be varied separately, or where their variations can be statistically distinguished by an analysis of variance.

I. *Environmental Conditions.* A. *In the School.* In dealing with school conditions it is comparatively easy for a psychologist holding an official position in the school inspectorate to check his inferences by watching the effects of different methods of treatment. The evidence gathered in this way leaves little doubt that in many cases the main, if not the sole, cause of maladjustment arises out of current conditions at the child's school. This is confirmed by the figures in the table.

In many of the cases there noted the pupil's own mental characteristics proved to be perfectly normal. Thus a child of average ability working in a class or school attended by pupils whose general ability is much higher than his, or again a bright child attending a school in a poor neighbourhood where the intellectual level is decidedly low, may easily become maladjusted. Judged by the percentages, the condition most frequently reported was the assignment of the child to a class where the work is too hard ; [1] yet the size of the correlation coefficient suggests that its influence may not be so important as might be inferred. More often it is the personal relations between the child, his teacher, or his school fellows that provide the chief aggravating factors. Among these cases the commonest are those in which the child's class teacher is in some way uncongenial, unsympathetic, or lacking in understanding. The heading ' uncongenial pupils ' refers mainly to cases of teasing, bullying, or snobbishness, usually on the ground of social or physical peculiarities. Absence from school, whether as a cause or as an effect of maladjustment, is the third most commonly reported condition. Change of school or department plays a significant part both with older and younger children : with older children the commonest type of case is that of the scholarship winner from a relatively uncultured home, who was the star pupil in the primary school, but turns out to be a backward pupil in the secondary or grammar school to which he has been sent ; with younger children maladjustment not infrequently follows promotion from the infants' school to the junior : such a change is very liable to produce maladjustment when the methods of teaching are abruptly altered, particularly in the case of those who are dull or suffer from special disabilities. It should be observed

[1] Three cases in our series (0·8 per cent.) were children whose ability was slightly above average, and who were being urged to work for a junior county scholarship. We believe that in the past both the frequency and the gravity of such cases has been much exaggerated. In the chapter on ' maladjustments in school ' included in the child guidance council's *Survey*, already quoted, nearly half the section on ' mental factors ' is devoted to this type of case : " an unduly high percentage of maladjustment," we are told, " centres round the minor scholarship examination." We are not, however, given figures to indicate precisely what the percentage is. Similar fears have been expressed, especially by family doctors, in their reports on some of our cases : in the majority of instances there appeared to be little ground for the view adopted. In the few cases where it was justified, the pressure came more often from the ambitious parent than from the school.

TABLE IV. FREQUENCY OF CAUSES

CONDITION	Maladjusted Group Per cent.	Control Group Per cent.	Correlation
I. ENVIRONMENTAL			
A. IN THE SCHOOL			
Repeated Absence from School	7·3	0·2	·19
Changes of School	6·8	2·4	·11
Inappropriate School	5·7	1·9	·10
Inappropriate Class :			
Work too Hard	16·4	6·0	·17
Work too Easy	5·7	1·5	·12
Uncongenial Teacher	12·7	1·2	·23
Uncongenial Pupils	8·1	0·0	·21
B. IN THE HOME			
1. Poverty and its Concomitants			
Very Poor	13·3	8·2	·08
Poor	32·7	22·3	·12
Comfortable	46·5	52·3	—·06
Well-to-do	9·3	18·2	—·14
Overcrowding	26·7	15·7	·13
Lack of Facilities for Recreation	19·5	6·3	·20
2. Defective Family Relations			
Father Dead or Absent	12·8	4·4	·15
Father Deserted, Separated, or Divorced	7·3	0·5	·17
Mother Dead	15·3	3·1	·23
Mother Deserted, Separated, or Divorced	7·5	0·2	·19
Mother out at Work	12·7	4·1	·16
Prolonged Absence from Parents	8·3	2·2	·13
Step- or Foster-father	9·3	4·1	·11
Step- or Foster-mother	15·1	1·9	·24
Illegitimate	8·6	0·2	·21
Only Child	10·9	3·6	·14
Defective Pre-school Conditions	24·2	16·7	·09
3. Defective Discipline, etc.			
Over-strict	17·7	2·2	·26
Weak or Vacillating	14·3	4·1	·18
Lack of Affection	23·9	1·5	·34
Disagreement about Control	4·2	0·0	·14
Parental Disagreement	5·1	1·2	·12
Disagreement with Sibs ; Jealousy	7·9	3·9	·08
Neurotic Parents	3·6	1·5	·07
4. Vicious Home			
Sexual Immorality	3·6	0·2	·13
Drunkenness	9·6	0·5	·22
Quarrelling	3·1	0·0	·12
C. OUT OF SCHOOL AND HOME			
Excessive Facilities for Amusement	8·8	7·3	·04
Defective Facilities for Amusement	9·6	5·1	·07
Undesirable Companions	5·3	1·2	·11
Lack of Companionship	5·3	2·4	·08
II. PHYSICAL			
A. DEVELOPMENTAL			
Under-developed	5·5	1·7	·11
Over-developed	7·8	1·5	·16
Noticeable Peculiarities	8·3	2·2	·14

B. PATHOLOGICAL			
General Ill-health	13·8	10·9	·04
Specific Ailments (current)	7·3	6·8	·01
Specific Ailments (past : excessive number of)	8·1	3·6	·09
Neurological Disorders	3·4	0·5	·12
Biochemical Disorders	5·6	2·4	·08
III. INTELLECTUAL			
A. Apparently INNATE			
1. General Intelligence			
(*a*) Mental Deficiency	2·4	1·2	·04
(*b*) Dullness	9·3	3·4	·13
(*c*) Superior	6·2	0·0	·17
2. Special Disabilities	5·3	0·5	·14
B. Predominantly ACQUIRED			
1. General Educational Backwardness	37·1	3·4	·42
2. Special Educational Backwardness			
(*a*) Reading	9·1	2·4	·13
(*b*) Spelling	2·1	0·0	·09
(*c*) Arithmetic	7·8	4·1	·07
IV. EMOTIONAL			
A. Apparently INNATE			
1. General Emotionality			
(*a*) Unstable	48·3	6·5	·47
(*b*) Temperamentally Defective	2·1	0·0	·09
2. Temperamental Tendency			
(*a*) Excessive Extraversion	23·6	2·2	·33
(*b*) Excessive Introversion	21·3	3·4	·27
3. Specific Emotions			
(*a*) Timidity	3·1	0·0	·12
(*b*) Anger	2·1	1·0	·04
(*c*) Sex	4·8	1·9	·07
(*d*) Wandering	2·9	0·0	·12
B. Predominantly ACQUIRED			
1. Developmental or Normal			
Adolescent Instability [1]	8·3	0·0	·21
Personal Sentiments	2·6	1·0	·08
Impersonal Sentiments	2·3	0·0	·10
2. Psychoneuroses			
Hysteria (chiefly Conversion)	1·8	0·0	·07
Obsessions and Compulsions	0·5	0·0	·03
Anxiety States	10·3	2·2	·16
Neurasthenic States	4·8	1·2	·11
Psychopathic or Pre-psychotic	2·9	0·0	·12
C. Habitual Emotional REACTIONS [2]			
Conflict (including Ambivalence)	7·1	0·0	·18
Regression	4·6	1·5	·09
Perseveration	2·3	0·0	·12
Feelings of Inferiority	9·3	1·2	·17
Feelings of Superiority	2·1	0·0	·10
Feelings of Insecurity	10·1	2·7	·15
Irritability	2·6	2·2	·03
Jealousy	4·4	3·6	·02
Miscellaneous Grievances	6·5	2·4	·11
Incontinence	4·1	1·2	·09
V. MORAL [3]			
1. Moral Conflicts	3·4	0·0	·12
2. Religious Conflicts	1·5	0·0	·08

[1] Mostly pupils over 14.
[2] These conditions are entered here only when they appear to be contributory causes to the maladjustment, not merely its effects.
[3] Cases referred for known delinquency have been excluded from this series.

that a significant correlation also arises in the case of bright pupils who have been left in classes or schools where the work is too easy and so affords insufficient outlet for their abilities or ambitions.

The importance we are led to attach to school conditions in so many cases of maladjustment runs counter to the view that has recently been put forward in several quarters. "The *primary* causes of maladjustment," it is said, "are to be found within the individual and in the structure of his relations with the rest of his family, especially as they operate during the first five years": when conditions in the school *appear* to produce maladjustment, what has really happened is that "a hidden disturbance has become manifest in response to school stresses." Now in following up the after-histories of the earlier cases, we have found that, in quite a high proportion, remedies directed solely to an alteration of the school conditions have been followed by a complete and apparently permanent disappearance of every overt sign of maladjustment. On this point the records kept by one of us over a period of nearly 20 years furnish ample evidence: thus, in 124 instances where a change of school was arranged, reports were available regarding the child's subsequent progress, and in 73 per cent. a complete improvement seems to have occurred; in 186 cases a change of class was carried out, and a similar improvement occurred in 68 per cent.; in many of these the main factor appears to have been a change to a more congenial teacher. Often, where special disabilities are concerned, simply altering the teaching methods has proved almost equally effective. This was conspicuously the case with young children suffering from a disability in reading that appeared to have provoked a widespread emotional reaction: here, out of 54 cases studied at different times, a satisfactory result was obtained in 35, i.e., in nearly two-thirds.

B. *In the Home.* In the reports on home circumstances a large number of different conditions were noted, and were frequently put forward as the main or sole cause. But once again the mere frequency with which such conditions are found is by no means a safe guide to their actual influence.

Poverty is a common characteristic among maladjusted children of the elementary school type; but the correlation is not high. A study of the individual cases falling under this head suggests that poverty only becomes a contributing factor when it tends to limit the natural development of the child's intellectual capacities or emotional interests. Thus an exceptionally bright child coming from a very poor home where there are few opportunities for indulging its intellectual interests, a lively and adventurous type of child brought up in a restricted slum environment, the child from a poor but highly respectable home sent by its parents to a school attended by wealthier children—all are apt to develop symptoms of maladjustment.

In the main the psychological conditions of the home are much more important than the economic or material. Our data fully bear out Sully's contention: the mental and moral influence of the home on the adjustments made by the child is "far greater than that of any other single factor," for four main reasons: (i) "first, its influence is far more comprehensive than that of school or other external circumstances"; (ii) "it acts much earlier, at a time when first impressions are formed"; (iii) its effects are far more sustained; (iv) finally, "family relations, involving warmer relations and more intimate intercourse, give a peculiar intensity to the influence of the home" ((16) p. 572).

Defective family relations play much the same part in maladjustment as they do in the causation of delinquency.[1] The absence of one or both parents, or the removal of a child from its parents for an appreciable period, is often followed by maladjustment, though, as a rule, such causes appear to be contributory rather than primary. The condition that has the most direct and the most powerful influence upon the child's adjustments to its daily life at home is the attitude of its parents or of those who act *in loco parentis*. From our experience it would be easy to classify the parents of such children into the four suggestive groups, somewhat picturesquely described by Lafore, namely, (i) 'dictators,' (ii) 'appeasers,' (iii) 'temporizers,' and (iv) 'co-operators' (10). Contrary to what is found in the investigation of delinquency, it is the 'dictators,' with their excessively strict disciplinary treatment, who seem to be responsible for the most serious cases of maladjustment: but the weak or vacillating discipline of the 'appeasers' and the 'temporizers' may also have effects that the casual visitor may scarcely suspect. The largest coefficient of all is associated with what can only be described as a general lack of affection. This includes a few instances of grave neglect (2·7 per cent.) or even downright cruelty (0·5 per cent.); but the majority are cases of mere indifference or selfishness.

Out of the long list of conditions reported as characterizing the child's home circumstances, seven stand out as yielding high correlations (coefficients over 0·20), namely, lack of affection, over-strict discipline, the presence of a step- or foster-mother, the death of the child's own mother, drunkenness, illegitimacy, and lack of adequate facilities for recreation. Several characteristics frequently stressed

[1] It has of late been the fashion to lay special stress on 'broken homes' and similar conditions as regular causes of maladjustment and delinquency. But the low figures in our table do not support this view. We may note that, in their recent study of *Patterns of Marriage*, Eliot Slater and Moya Woodside found that broken homes were as common in the early lives of the well adjusted as in those of the neurotic. In *The American Soldier* Dr. Stouffer and his colleagues come to the same conclusion.

in the literature appear to have less importance than is commonly supposed ; for instance, prolonged absence from parents during earlier childhood, disagreement between the parents in regard to the child's own management or other affairs, and the fact that one or other parent (usually the mother) is highly neurotic. So far as possible, we obtained detailed information regarding the child's early history before reaching school age. For this we relied, not only on the mother's recollection, but on reports from teachers, social workers, and others, who had often been acquainted with the household for over a long period of time. Among the maladjusted group there were a large number of cases in which material and emotional deprivations were reported, together with stories of early difficulties and injudicious handling in regard to such problems as weaning, feeding, cleanliness and the like. But in the control group (where our information is probably less complete) a high proportion of such cases was also reported. Hence, as will be seen from Table IV, the coefficient of correlation is barely significant.[1]

The environment outside the home and the school plays a smaller part here than it does in cases of juvenile delinquency or among those who have already left school.

II. *Personal Characteristics.* A. *Physical Conditions.* Unless they affect the nervous system, physical weakness and disease appear to have comparatively little direct influence. Where change of school involves a change in social or cultural environment (e.g., the weakly scholarship winner sent to a secondary or residential public school), physical ill health and lack of bodily strength may prove important ; but our list contains only a few instances of this type, and most of them should perhaps be primarily regarded as cases of an unsuitable school. Irregular development in physical characteristics, particularly among girls, may occasionally prove a temporary factor ; and the presence of some conspicuous physical peculiarity—a squint, a stammer, a cleft palate, left-handedness, or even red hair—may indirectly lead to maladjustment, though as a rule such cases are referred on other grounds. If our data can be trusted, therefore, the importance of physical conditions and pathological defects has been greatly exaggerated.

The only significant correlation is that for neurological disorders, which are however decidedly rare. Under this heading the conditions noted were similar to those reported in the investigations on delinquency ((2) p. 250) ; by far the commonest were post-encephalitic conditions, migrainous headaches, and chorea (usually associated with rheumatism) : we have also included three cases in which the maladjustment appeared to be due almost exclusively to partial deafness. Under biochemical disorders the majority of cases consisted of disturbances of endocrine functions and of carbohydrate metabolism, together with a few alleged cases of vitamin deficiency and intolerance of fat : so far as our evidence goes, it seems clear that the importance of such conditions cannot be very great.[2]

B. *Intellectual Characteristics.* Our case-histories leave little doubt that innate disabilities, both general and special, may form a predominant factor in the production of maladjustment. But from the table it appears that the correlation of maladjustment with acquired educational backwardness is far higher than its correlation with disabilities that are presumably innate : the figure, indeed, is one of the highest in our list. However, judged by the after-histories, the condition would seem to be quite as often an effect of the maladjustment as its cause.

[1] We cannot help thinking that the emphasis placed on early conditions in psychiatric writings is due partly to unverified theories based on psychoanalytic studies of adults, partly on the fact that the psychiatric specialist is far more familiar with early conditions in middle-class and well-to-do homes than he is with conditions among the general elementary school population. Methods of parental treatment, such as are quite common in large families in poorer neighbourhoods, would seem positively shocking if found, say, in a family of the professional classes. At the same time we have no wish to deny the importance of home treatment during the first few years of life, particularly in its effects on the formation of acquired emotional and social characteristics : in the case of delinquents, for instance, the correlation, though not high, is unquestionably significant.

[2] In three cases where ' acidosis ' or ' glycopenia ' was reported, the alleged maladjustment cleared up after appropriate treatment. But, in the interests of the children, the treatment always included a number of different measures, and the inferences to be drawn were never very clear. As regards endocrine disturbances our results are fully in keeping with the views expressed by Professor Moncrieff : " sweeping and unscientific claims for endocrinal disorders as factors in mental disturbances are not supported by satisfactory proof : clear-cut examples occur in the literature ; but in general, therapy based on speculative relations of this type is futile " ((5) p. 63).

As we have seen, in the present series most of the cases falling under this head were not, in point of fact, referred primarily on the ground that the child was dull or backward. Often this was discovered only incidentally. Indeed, in many cases the amount of retardation was too slight for the child to be regarded as dull or backward in the technical sense, or (as it would now be called) 'educationally subnormal.' It may be noted that an innate degree of general intelligence *above* the general average is apparently a more effective (though not a more frequent) cause than an innate degree *below* the general average. Often the difficulty can be remedied by a change of school or class. But quite as frequently it arises out of school and in the home : the boy who is much brighter than his parents, or than his brothers and sisters, sometimes presents an exceptionally puzzling problem ; and it is one which seems rarely to be noted in the literature.

C. *Emotional Characteristics.* If we may judge by the total percentages, emotional conditions appear to be far more important than any that we have considered hitherto. Their numerical predominance in the present series, however, is due partly to the fact that cases referred specifically on the grounds of intellectual subnormality or delinquency were not included. In the cases falling under this general heading, the most conspicuous feature is the presence of a high general emotionality, producing what is sometimes termed an unstable temperament. This furnishes the largest correlation in the whole list.

Among a few of the older children the instability proved to be a temporary condition, incidental to puberty or adolescence ; but in the vast majority it appeared to be innate.[1] Often other members of the household, particularly the parents, were equally unstable, so that environmental causes were operative as well.

Among the cases referred by teachers, the extraverted, aggressive, self-assertive type was the commonest ; among those referred by school medical officers or discovered in the course of school surveys, it was rather the introverted child, suffering his deprivations and discomforts in silence, and appearing worried or unhappy for no obvious cause, that was most frequently noted. Both these more extreme types seem to find it difficult to make spontaneous readjustments to the shortcomings and frustrations which at one time or another confront so many children in homes of limited means.

In an appreciable proportion of the cases, the maladjustment proved to be the cumulative outcome of unfavourable or uncongenial conditions that had lasted for a period of years, and had thus gradually built up undesirable but habitual emotional reactions—continual mental conflicts or a deep sense of inferiority or insecurity. The highest correlation in this subgroup is found in the case of mental conflict : (cf. (3) pp. 206f.).[2] A study of the individual cases, however, suggests that, even when conflict appears as a cause rather than as an effect, it nevertheless remains a contributory factor only : it is chiefly when repeated or prolonged conflict is associated with material or psychological restrictions imposed by the environment that symptoms of maladjustment are likely to ensue.[3]

[1] Those who have criticized the hypothesis of innate differences in general emotionality generally do so on the ground that such differences, though appearing early, are attributable to the treatment of the child during the first year or two of life : in particular, it is alleged, " an emotional mother is apt to produce an emotional child by her postnatal treatment of the infant rather than by her transmission of an hereditary trait." But this view does not explain the existence of wide differences among siblings treated in the same fashion, nor the high and significant correlations that can be demonstrated between children and remoter relatives. The fact that, by selective breeding, it is possible within eight generations to produce distinct groups of rats—one highly emotional and the other markedly unemotional, as shown in situations causing conflict or other forms of maladjustment—appears to confirm our hypothesis : (cf., for example, Hall, C., *Sigma Xi Quarterly*, XXVI, 1938, pp. 17f.).

[2] Since it has been so frequently stated that " the part played by conflict in the production of delinquency and emotional maladjustment was ignored by academic psychology, and its recognition is due to Freud," may we point out that both conflict and the principle of 'sublimation' (though the metaphor was not used) are discussed by Sully in his *Handbook* (pp. 542f.) and by Stout in his *Analytic Psychology* (I, pp. 281f., and II, pp. 148f.) ?

In addition to those forms described elsewhere in discussing conflict as a causal factor in delinquency ((2) pp. 553f.) we may note one type that seemed particularly common in our groups, namely, that arising from incompatible ideals at school and home respectively : among his school fellows, for example, the ideal youngster is a bold, aggressive, adventurous creature ; with his parents he is a quiet, submissive child ; and the need to hold one's own by self-assertion in one environment and the equally powerful need to retain approval and affection in the other environment often lead to conflict and maladjustment. In our experience this is quite as frequent as the conflicts arising out of sex, repressed anxiety, or supposed insecurity, which are so repeatedly stressed in the literature. Piaget has stressed a similar point in his discussion of group loyalties.

[3] This is in keeping with recent work on the production of experimental neuroses in animals. With them, conflict appears to be a necessary, but not a sufficient factor in the causation of behaviour disorders. Conflict must be accompanied by restriction, either physical (as with Pavlov's harnessed dogs) or mental (imposed by the animal's own limitations or fixed habits) : cf. Liddell, H. S., and Anderson, L., *Arch. Neurol. Psychiat.*, XXXIV, 1935, pp. 330f.

The high percentage given for 'feelings of insecurity' is perhaps due to the fact that we have here grouped together rather a wide variety of different reactions, based chiefly on statements or impressions obtained during interviews with the children: they have not been counted as potential causes of maladjustment unless they amounted to fairly persistent states of anxiety (as a rule by no means baseless). The percentage for the normal children is probably too low, since they were not studied so intensively.[1] In 14 per cent. of the maladjusted cases these emotional reactions took the form of symptoms that would be generally regarded as definitely pathological or abnormal, so that the child has been diagnosed as psychoneurotic or (in 3 per cent.) as psychotic.[2]

We should add that it is clear from our data that the incidence and relative importance of the various conditions differ appreciably in different districts, among different social classes, and, in a lesser degree, among different age-groups. But our total groups are too small for these variations to be studied in detail. We venture to suggest that the next most fruitful topic for research might well be the differences between the causes chiefly operative (*a*) at different ages and (*b*) in different social groups.

VI. SUMMARY AND CONCLUSIONS

1. The correlations between 24 conditions reported among 273 maladjusted children have been factorized. Only a small general factor was found. The first bipolar factor indicates a broad classification into environmental conditions and personal conditions respectively; the second, a cross-classification into intellectual conditions and emotional. The group factor analysis suggests a corresponding subdivision into four distinguishable types; these, however, show considerable overlapping. It is concluded that cases of maladjustment can hardly be regarded as forming a single relatively homogeneous group.

2. The frequencies of the various conditions have been examined in fuller detail among a series of 394 cases of maladjusted children, and have been compared with those obtaining among a control group consisting of the same number of normal children. Point-distribution correlations have been calculated from the percentages found. Judged by these coefficients, a wide variety of conditions shows significant correlations with maladjustment. The most important appear to be the following: (i) in the school, the presence of an uncongenial teacher, and assignment to a class where the work is too difficult; (ii) in the home, lack of affection, over-strict discipline, presence of a step- or foster-mother, death of the child's own mother, drunken parents, illegitimacy, and lack of adequate facilities for recreation; (iii) as regards intellectual characteristics, general educational backwardness, often (though not always) the result of innate intellectual disability, either general or specific; (iv) as regards emotional characteristics, general emotional instability, excessive extraversion, excessive introversion (often accompanied by anxiety states, conflicts, or feelings of inferiority), and (among the older children) adolescent instability. There are appreciable variations according to age, locality, and social class.

[1] In this insecure age a good deal of stress is laid on the 'need for security' which children are presumed to show: the phrase, however, covers a wide variety of conditions; and moreover we venture to think that, particularly in a civilized society, the need for the excitement and the sense of adventure furnished by conditions of insecurity is often as great.

[2] A very large proportion of our cases would probably have been regarded as suffering from psychoneurotic disorders by many psychiatrists, and a number (including most of those in the clinic groups) had actually been described as 'mentally ill.' But, when the child's behaviour is a perfectly normal reaction to an abnormal situation, and when all that is needed is a readjustment in the social environment (e.g., at home or at school), we consider it misleading to term the condition an 'illness.'

REFERENCES

1. Allport, G. W. (1938). *Personality : A Psychological Interpretation.* London : Constable.
2. Burt, C. (1925). *The Young Delinquent.* London : The University of London Press.
3. Burt, C. (1933). *The Subnormal Mind.* London : Oxford University Press.
4. Burt, C. (1946). ' The Assessment of Personality.' *Egypt. J. Psychol.*, II, 1–21.
5. Gordon, R. G., ed. (1939). *A Survey of Child Psychiatry.* London : Oxford University Press.
6. Hall, M. B. (1947). *Psychiatric Examination of the School Child.* London : Arnold.
7. Harms, E., ed. (1947). *Handbook of Child Guidance.* New York City : Child Care Publications.
8. Harrowes, W. (1949). *Human Personality and its Minor Disorders.* Edinburgh : Livingstone.
9. Henderson, D. K., and Gillespie, R. D. (1933). *Textbook of Psychiatry.* London : Oxford University Press.
10. Lafore, G. G. (1945). ' The Practices of Parents in Dealing with Pre-school Children,' *Child Development Monographs*, No. 31. New York : Teachers' College, Columbia University.
11. Meyer, A. (1935). ' The Psychobiological Point of View,' *ap.* National Research Council, *The Problem of Mental Disorder.* New York : McGraw-Hill Book Co.
12. Ministry of Education (1945). *Handicapped Pupils and School Health Service Regulations.* London : H.M. Stationery Office.
13. Rogers, C. R. (1942). ' A Study of Mental Health Problems in Three Representative Elementary Schools,' *Bureau of Educational Research Monographs*, No. 25. Ohio : Ohio State University.
14. Sandiford, P. (1938). *Foundations of Educational Psychology.* London : Longmans, Green, and Co.
15. Shryock, R. H. (1948). *The Development of Modern Medicine : An Interpretation of the Social and Scientific Factors Involved.* London : Victor Gollancz.
16. Sully, J. (1894). *Teachers' Handbook of Psychology.* London : Longmans, Green, and Co.

APPENDIX : THE POINT CORRELATION WITH EQUAL NUMBERS

When an experimental group has been matched with a control group, and percentages have been calculated showing the relative frequencies with which a series of conditions have been found in each, a convenient method of assessing the association between any one condition and the basic feature used to distinguish the two groups is provided by a product-moment coefficient computed on the assumptions (*a*) that the numbers in the groups are equal and (*b*) that the conditions conform to a point distribution. With these assumptions, the correlation (or ' phi coefficient ') is

$$\varphi = \frac{p_e - p_c}{\sqrt{pq}}, \qquad \text{(i)}$$

where p_e and p_c denote the proportions found in the experimental and control groups respectively, p is the total of these two proportions, i.e., $p_e + p_c$, and $q = 1 - p$.

When a large number of such correlations have to be calculated, and the requisite computing tables are available, a quicker working procedure is provided by changing the denominator to terms of the *average* proportions : i.e., by taking

$$\varphi = \frac{p_c - p_e}{2\sqrt{\bar{p}\bar{q}}}, \qquad \text{(ii)}$$

where $\bar{p}$ denotes the average, i.e., $\frac{1}{2}(p_e + p_c)$. Values for $\sqrt{pq}$ can be found from appendices to various statistical text-books or in Table XV in Holzinger's *Statistical Tables*. Thus, if in the maladjusted group the percentage of children whose mothers are working during the day is 24 per cent., and in the control group only 16 per cent., then $\bar{p} = \frac{1}{2}(24 + 16) = 20$, and

$$\varphi = \frac{\cdot 24 - \cdot 16}{2\sqrt{\cdot 20 \times \cdot 80}} = \frac{\cdot 08}{\cdot 80} = \cdot 10 .$$

In such cases time is saved if a nomogram or ' abac ' is plotted for convenient values similar to that shown in Fig. 27 of *Mental and Scholastic Tests* (1921 ed., p. 219). The co-ordinates are readily constructed by putting equation (ii) in the form

$$p_e = \bar{p} + \varphi\sqrt{\bar{p}\bar{q}} \text{ and } p_c = \bar{p}\ \varphi - \sqrt{\bar{p}\bar{q}} \qquad \text{(iii)}$$

and taking $\varphi = \cdot 05, \cdot 10, \cdot 15$, etc. When the average percentage is 50 per cent., it is clear from equation (ii) that the correlation will be equal to the simple difference between the two proportions. When the average percentage diminishes while the difference remains the same, the value of the correlation increases, in inverse ratio to the standard deviation ($\sqrt{pq}$). Thus proportions of 12 per cent. and 2 per cent. furnish a correlation twice as large as that furnished by proportions of 52 and 42 per cent., although the difference is the same.

Where the experimental group forms the tail end of a normally distributed population, cut off sharply from the control group, and where the conditions to be compared are themselves based on normally distributed characteristics, a tetrachoric coefficient would seem more appropriate. But,

where the form of the distribution for the classifying attributes (here presence or absence of maladjustment) appears uncertain, where the method of subdivision is not clear cut, and where many of the conditions to be compared turn upon the presence or absence of some more specific feature, there the procedure described above will provide the simplest and most satisfactory mode of comparison.[1]

CORRESPONDENCE

(1) THE ESTIMATION OF COMMUNALITIES

To the Editors, British Journal of Psychology : Statistical Section.

Sirs,—In his reply to Professor Thomson (this *Journal*, III, p. 127) Professor Burt states that his " reduction of self-correlations is based, not, as Thomson seems to suppose, on considerations of rank, but on considerations of significance." Elsewhere he has criticized most of the familiar rules put forward by various writers for estimating ' communalities ' or ' self-correlations ' ; but it is difficult to find any explicit statement of the rule he himself would recommend. In the early *Note* on working procedures (1934) to which he refers (in this *Journal*, II, p. 60) the following correlation table is used to illustrate both ' weighted summation ' and ' correction for specificity.' The matrix of observed correlations has a minimum rank of 3. Owing to the large size of the sample, it is assumed that all three factors are fully significant ; and this of course leads to perfectly definite determinations of the self-correlations (as shown by the figures in brackets). But suppose that only one factor is significant : what self-correlations would he suggest in carrying out the two alternative methods of factorization proposed ?

Tests	1	2	3	4	5	6
1	(·82)	·42	·20	·12	·13	—·10
2	·42	(·46)	·12	·22	·23	·08
3	·20	·12	(·56)	·20	—·10	·24
4	·12	·22	·20	(·17)	·08	·15
5	·13	·23	—·10	·08	(·17)	·00
6	—·10	·08	·24	·15	·00	(·21)

A. L. DIXON.

Reply.—In my view there is no *single* rule for mechanically determining the most appropriate values for the self-correlations. Everything depends on the nature of the question to be answered.

If the investigator's hypothesis is (i) that all except the first r factors are definitely non-significant, and (ii) that the residual ' errors ' in the observed correlations are entirely independent, then the self-correlations will be sums of the squares of the saturation coefficients for the r significant common factors. In general, therefore, they will have to be ascertained by a process of successive approximation. In the case of the table quoted, the saturations for a single factor would (with the foregoing assumptions) have the values shown below, and the self-correlations would be the squares of these figures.[2] It follows that, no matter which method of factorization is employed, the matrix of residual correlations will contain zeros in the leading diagonal ; and (unless we reintroduce part of the excluded factor variance) it will no longer be Gramian. However, the columns (or rows) of residuals will still fulfil the familiar requirements : with simple summation they will add up to zero as they stand ; with weighted summation (or with correction for specificity) they will add up to zero when they have been weighted in the appropriate way.

[1] For further discussion see *Mental and Scholastic Tests, loc. cit. sup.*, *The Backward Child*, Appendix III (and refs.), and Yule, U., ' Methods of Measuring the Association between Two Attributes,' *J. Roy. Stat. Soc.*, LXXV, 1912, 579–642.

[2] These figures were given in an earlier version of the *Notes* to which Mr. Dixon refers. The data and number in the sample were imaginary, and used merely to illustrate certain technical points.

This seems a fairly satisfactory principle to adopt when we wish to subject some clearly defined hypothesis to a rigorous test : (it was, for instance, used in early discussions of Spearman's single-factor hypothesis). But for practical purposes it is decidedly laborious. Further, the errors of the observed coefficients are not strictly independent ; and in psychological work we can nearly always be sure of the presence of small common factors in addition to those that are fully significant. As a rule, therefore, the computer can be content with approximate determinations of the self-correlations, and these should, if anything, err on the side of a small excess.

SATURATIONS FOR GENERAL FACTOR

Tests	1	2	3	4	5	6
By Weighted Summation ..	·535	·741	·262	·325	·235	·109
With Correction for Specificity ..	·532	·783	·206	·290	·266	·086

C. B.

(2) TETRACHORIC CORRELATIONS AND DIFFICULTY FACTORS

SIRS,—I wish to make a correction to a footnote in my article (this *Journal*, IV, 1951, p. 72). The clause " and this in turn would disturb the normality of the bivariate distribution " should not have appeared. The expression is meaningless in the context where it is used. I understand that Professor Burt and Miss John did use some correction procedure for dealing with the case of a zero frequency in a fourfold table ; but I still cannot myself see any justification for such a procedure. As I stated in my footnote, the tetrachoric correlation in such cases is unity.—NEIL GOURLAY.

SIRS,—In their reply to Mr. Gourlay's criticisms of their work on the Binet tests, Miss John and Professor Burt suggest that it would have been better to substitute ' point correlations ' for ' tetrachoric ' (this *Journal*, IV, p. 75). This presumably refers to Yule's form of the phi-coefficient used by Professor Burt in his analysis of the results obtained with such tests (*Mental and Scholastic Tests*, 1921, ed., pp. 217f.). But surely if (as in the case referred to by Mr. Gourlay in his table) " one of the two items has a frequency of zero in one cell," then the value of the correlation should still be unity. The coefficient is definitely stated by Yule himself to have the value of $+1$ when one of the frequencies is zero (*Introduction to the Theory of Statistics*, 1912, p. 39).—E. N. WARNER.

Reply. The coefficient proposed was not the *equalized* phi-coefficient, as Mr. Warner supposes (elsewhere termed by Yule the ' coefficient of colligation ' and so described in *Mental and Scholastic Tests*), but the more general phi-coefficient discussed by him on p. 217 of the volume which Mr. Warner cites. We assumed this would have been clear from the formula we gave on a later page (eq. (i)). In Mr. Gourlay's notation, Yule's eq. (13) may be written

$$\varphi = \frac{ac - bd}{\sqrt{\{(a+b)(b+c)(c+d)(d+a)\}}} .$$

When, as in the table referred to, $d = 0$, this becomes

$$\frac{ac}{\sqrt{\{ac(a+b)(b+c)\}}} = \sqrt{\frac{ac}{(a+b)(b+c)}} .$$

If, however, the subtotals are equalized, $a + b = b$ and $a + d = d$. Consequently $a = d = 0$; and in that case $\varphi = +1$ (as Mr. Warner asserts). But, with the vocabulary tests, the subtotals cannot be treated as equal, since by hypothesis one test is ' more difficult ' than the other, i.e., answered correctly by a smaller proportion.

With the conditions assumed by Mr. Gourlay, the tetrachoric correlation should no doubt be unity. But in that case the distribution ceases to be bivariate. We believe that the difficulty arises from confusing the correlation between what we might call the manifest variates and the latent variates—between the measurements and the things measured. The measurements (which take the form of frequencies) obey a point-distribution, and yield a misleading result if treated by the tetrachoric procedure ; the underlying processes, however, are assumed to have a normal distribution, and we consider it equally misleading to assume on *a priori* grounds that they must always be in perfect correlation just because the items from one test are included in the other.

ENID JOHN.
CYRIL BURT.

BOOK REVIEWS

An Introduction to Modern Psychology. By O. L. Zangwill. London: Methuen & Co. Pp. xii + 227. 5*s.* net.

This volume is the latest addition to the admirable series described as *Home Study Books*. Its purpose, we are told, "is to provide an outline of modern psychology as envisaged from the biological standpoint": "trends in present-day psychology are assessed with a view to their bearing on the psychology of to-morrow." The chief subjects selected for discussion are experimental psychology, physiological psychology, psychoanalysis, and the various problems connected with intelligence and mental measurement. On the other hand, a number of topics, which are of more popular interest, and at the same time characteristic of contemporary psychology, are somewhat summarily dismissed. As the author observes, "but little attention is devoted to the problems of personality and none whatever to the study of psychology in its social aspects; the former," he believes, "is still in its infancy, and the latter remains a dream rather than an accepted discipline."

In his assessment of various 'present-day trends,' Mr. Zangwill bestows his highest praise on the current tendency to interpret mental life in terms of 'organic neurology,' and condemns as a misguided failure the whole attempt to introduce quantitative methods and statistical analysis into psychological investigation. "Mathematical psychologists," he says (quoting an earlier judgement by Dr. Chambers), "build dizzy numerical edifices, forgetting in their numerical zeal the flimsy foundations upon which their fabrics stand"; and he supports this verdict by a detailed criticism of the work carried out with mental tests during the last half century. One may perhaps doubt whether an elementary manual for the general reader is altogether an appropriate place for such a highly technical discussion. But, since Mr. Zangwill has developed his views at some length, and since (as his quotations show) there are other psychologists—and many non-psychologists—who share his scepticism, it seems desirable to examine the grounds on which his objections to statistical methods are based.

Mr. Zangwill confesses that his standpoint "admittedly reflects bias." And certainly his statements do not appear altogether consistent. But his main argument would seem to run as follows. The foundation of all 'mathematical psychology' is the assumption that 'psychological qualities can be measured.' Now Mr. Zangwill believes he can show that "nothing psychological is truly measurable." In the first place, "psychological qualities do not function in isolation," and to measure a thing you must isolate it. Secondly, "human capacity is too complex for one to place any great reliance on psychometric ratings," since what is complex cannot be ranged along a linear scale: hence, it becomes "exceedingly doubtful" what significance, if any, can be attached to "marks or measurements assigned to the results of a single testing." And we are thus forced to conclude that "it is very doubtful whether the concept of measurable quality may be applied at all to psychological qualities."

All this seems rather sweeping. Mr. Zangwill's statements would be plausible if they were prefixed by the qualification "sometimes" or "in some cases"; but then of course his wholesale generalizations would no longer follow. In his chapter on experimental psychology he himself accepts the idea of measuring sensory thresholds: he even prints a diagram showing measurements of 'brightness of sensation produced by light flashes of different intensity in man.' Here surely is a 'psychological quality' which is 'measurable.' What Mr. Zangwill chiefly means to question, therefore, would seem to be the idea of measuring *individual differences* in these various qualities. As he says elsewhere, he thinks we can "dismiss the theory and practice of mental testing as mere technology," and baseless technology at that; it is baseless because, as he once again insists, no reliance can be placed on differential measurements obtained from different individuals at a single sitting. Thus, so long as the psychologists are content merely to measure changes in visual acuity as a general quality (with a view to showing how it depends on varying conditions of illumination and the like) Mr. Zangwill is prepared to admit the conclusions which they draw. But when, with precisely the same apparatus and procedure, they attempt to measure John Brown's visual acuity (with a view, say, to discovering whether John's bad reading may be due to his imperfect vision or whether his vision could be improved by spectacles of a suitable dioptric strength), Mr. Zangwill would at once object that, in the case of a single individual, it is hopeless to rely on the results of a single testing. But can we not improve the reliability of a single testing by repeating the test, and then averaging the results? This is a common remedy which Mr. Zangwill never actually mentions. No doubt he would reply that, since each single test by itself is valueless, a whole series of such tests must be equally devoid of value: by adding together a number of worthless links you cannot construct a worth-while chain.

This kind of criticism is by no means new in the history of youthful sciences. The arguments which Mr. Zangwill adduces against 'mathematical psychology' are very similar to those that have regularly been urged time after time by non-mathematical critics against the attempt to introduce

quantitative methods into most other branches of science. Swift caricatured Newton ; Davy criticized Lavoisier ; Rumsey protested against Chadwick's proposal to introduce vital statistics into medicine ; Cairnes and Sidgwick rejected Jevons's attempts to apply algebraic formulæ into economics ; and nearly all biologists of Karl Pearson's day criticized his biometric innovations. Certainly, if at the very outset we are forbidden to estimate the predictive value of ' a single testing ' by some precise numerical index based on probability, then that value must remain " exceedingly doubtful." But such arguments would tell most heavily against the use of statistics in those very fields for which it is most appropriate—namely, cases in which the observable phenomena are highly complex and the causal influences rarely function in actual isolation. Consider, for example, the various physiological and pathological tests that would nowadays be applied by a physician when examining an applicant for life assurance. Mr. Zangwill would presumably assure him that " physiological qualities do not function in isolation," that " human vitality is too complex for one to place any reliance on physiological ratings," and that it is " exceedingly doubtful what significance, if any, can be attached to the conclusions reached as a result of a single examination." But if such objections had the smallest weight, then all insurance societies should long ago have collapsed into bankruptcy.

Now, in point of fact, the conditions which render any quality, whether physical or psychological, amenable to measurement have been repeatedly laid down, in clear and systematic fashion, by recent writers on scientific methodology ; and the statistical psychologist has more than once offered closely reasoned arguments to show that many mental qualities fully satisfy the requisite conditions.[1] Mr. Zangwill never attempts to explain where he considers these arguments to be at fault. Could he perhaps be persuaded to do so in a contribution to this journal ? Meanwhile, unless forewarned, his readers may easily infer that the advocates of mental measurement never stop to examine what Mr. Zangwill describes as " the uncertain basis " on which their work rests.

It is more particularly for factor analysis that his severest strictures are reserved. " Factor analysis," he says, " proceeds first by designing arbitrary tests of abilities ; secondly, by assigning arbitrary scores to the results so obtained ; and thirdly, by submitting the resulting correlations to a mathematical analysis which results in the isolation of factors. These factors, whose psychological significance (if any) is unknown, are then identified with the original abilities." And he concludes that, " if these objections are valid, then, we may submit, with due humility, that factorial analysis, despite its impressive mathematical procedures, will strike the future historian of psychology as a brilliant but misguided departure from the central path of empirical psychology." No doubt, if factor analysis did proceed in this way, then, as Mr. Zangwill asserts, it would seem " distinctly doubtful whether such methods could ever provide a satisfactory classification of human aptitudes." But would it not have been fairer to present the factorist's own account of how he actually proceeds ?

It is true that, to explain the method in fuller detail, Mr. Zangwill thinks that " we can hardly do better than quote Spearman's own account." And this he does at some length. Spearman's approach, we are told, " was by applying the method of correlation to the data of mental tests " ; and for him the value of the correlation technique lay in providing a means for analysing the correlated data in terms of factors. Accordingly, " the first step in factorial analysis is to ascertain whether the tetrad equation fits any table of actually observed correlations." From this we are led on to Spearman's early theory of two factors, his views on the nature of *g*, and a critical discussion of the concept of ' general intelligence.' And the outcome of all this, so we are assured, has not been " any exceptional understanding of the nature of intelligence : indeed, for the vast majority of purposes, the psychologist allows his definition of the term ' intelligence ' to be governed by popular usage." Hence, in spite of fifty years' work, " intellectual testing is a technology whose foundations are distinctly insecure."

Now, if Mr. Zangwill had merely concluded that the foundations of *Spearman's* views on intellectual testing were " distinctly insecure," most psychologists would probably have agreed with him ; but they would then be tempted to ask—why trouble the ' home student ' with the ' theory of two factors,' now that it has little more than an historical interest even for the highly specialized student ? Moreover the account which Mr. Zangwill gives is decidedly unfair to Spearman's immediate predecessors. Nowhere do we find the slightest hint that the application of correlation techniques to test-results and other anthropometric data had been proposed several years before Spearman entered the field, and that the analysis of correlated data in terms of uncorrelated components or factors was first suggested, not by Spearman, but by what Spearman called the ' Galton-Pearson school.' Spearman's own procedure, which is not factor analysis as ordinarily understood, was put forward as a simplified modification of these earlier techniques ; and it issued in a divergent line of development which, as Mr. Zangwill himself points out in a footnote, " is not now accepted by the majority of statistical psychologists," and was, indeed, hardly ever employed except by Spearman's own students. To-day it is generally agreed that the outcome of the long controversy between the

[1] Cf., for example, Burt, C., *Factors of the Mind*, pp. 115f., or (for a more popular account) *id.*, ' The Theory of Mental Testing,' *Yearbook of Education*, 1950, pp. 52–8.

' Spearman school ' and the ' Galton-Pearson ' has gone almost entirely in favour of the older writers ; and so, as Mr. Zangwill puts it, " the Spearman factors will take their not unworthy place in the limbo of discarded elements of mind." Perhaps in a future edition he will add that the original methods developed by Pearson and his followers have not only survived the objections urged by Spearman, but have during recent years found fruitful application in other branches of science besides psychology.

The general picture which Mr. Zangwill gives of factorial methods has been put forward more than once by non-mathematical writers, who are fond of complaining (not quite fairly) that the factorists themselves never stop to explain the logical foundations of their methods. In spite of what he says, no factorist has ever supposed that " psychological qualities function in isolation." If they did function in that way, there would have been no need for statistical analysis ; experimental analysis could have taken its place. Nor does the ' psychometrist ' proceed by designing " arbitrary tests " and " assigning arbitrary scores." What then is the underlying logic of the factorist's experimental procedure ? He begins with a tentative hypothesis, based on previous observations—for instance, Galton's assumption that intellectual processes at all levels of mental life depend in varying degrees on a single innate general ability. Then, to test this hypothesis (re-formulated perhaps in more precise terms), the factorist ' designs tests ' to measure various abilities typical of each cognitive level ; and he assigns the scores, not in any ' arbitrary ' fashion, but on the assumption of an approximately normal distribution. Recognized statistical tests are then formally applied ; and the results, so it is inferred, fully confirm the initial hypothesis, and in addition reveal the presence of supplementary abilities of a more specialized type. Further correlational evidence shows this general factor to be largely innate. Finally, to avoid inventing pedantic terms, the word ' intelligence ' is adopted as a convenient label to designate this ' innate general cognitive factor.' Now does not all this correspond almost exactly[1] to " the principal stages " of a scientific enquiry, as described by Mr. Zangwill himself on p. 160 of his book ? And incidentally we now see that it is quite untrue to say that the definition of the term intelligence is " governed by popular usage " or that " psychologists are still not agreed as to its definition."

In his section on ' *Adult* Intelligence ' Mr. Zangwill says that " it is customary to class persons with an I.Q. of 70 or below as ' feeble-minded ' and those with an I.Q. between 70 and 80 as ' borderline defectives.' " This is hardly correct : Mr. Zangwill seems to have confused the borderlines for defective *adults* (as defined by the Mental Deficiency Act) with the borderlines for *children* certified for what used to be called ' special schools for the mental defective ' (under the old Education Act). In its discussion of adult defectives the *Report of the Mental Deficiency Committee of the Board of Education and Board of Control* (published over 20 years ago) suggests, so far as test-results are concerned, a borderline of 60 I.Q. (Pt. IV, p. 46). The higher borderline given by Mr. Zangwill would treble the proportion of certifiable adults.

In dealing with the problem of temperament and its relations to body-type, he discusses in some detail the theories of Kretschmer and Sheldon. But here no reference whatever is made to the more exact and cautious conclusions derived by factorial methods. Apparently Mr. Zangwill is quite prepared to accept such hypotheses provided they are *not* supported by exact statistical investigation, but rest solely on ' clinical observation ' and ingenious surmise. In short, as another reviewer has rightly observed, " Mr. Zangwill is decidedly unfair to the whole broad field of differential psychology, its foundations, techniques, and conclusions," and is " far more predisposed to accept the scientifically imperfect claims of clinicians and psychoanalysts than those of psychometricians." [2]

Mr. Zangwill is almost equally severe in his criticisms of the great names in experimental psychology. It was, we gather, a great misfortune that Fechner's early work " gave heart to all who believe that the data of psychology are open to mathematical treatment." And " the heritage of Wundt," so the ' home student ' is told, " is to-day scarcely recognizable, except perhaps in the preservation of antique brass instruments whose value is largely decorative."

Two detailed chapters give an admirable survey of ' modern physiological psychology.' Here the author is more at home. Yet it is doubtful whether recent psychological history would support his high assessment of the results of a purely neurological approach. It was the adoption of a ' purely neurological approach ' that led doctors to certify the merely backward as mentally defective, to diagnose backward readers as suffering from a ' congenital defect in the word-centre,' and to suggest surgical operations as a cure for mental deficiency. It was the introduction of intelligence-tests, and of the statistical analysis of actual cases, that overthrew these crude hypotheses. Mr. Zangwill refers to the abnormal electro-encephalograms ' observed in a number of problem children,' and infers that their " unsatisfactory conduct is primarily due, not to emotional maladjustment as is commonly supposed, but to constitutional defect at the highest levels of cerebral organization." Experienced psychologists who have actually worked with ' problem children ' would probably

[1] If Mr. Zangwill considers that this is not an accurate account of what factorists actually do, will he refer to the discussion of the ' general nature of the inductive reasoning on which factorial studies rest ' in *Factors of the Mind*, pp. 37f., or to Thurstone's later and briefer discussion of ' the role of hypotheses in factor analysis ' in *Multiple Factor Analysis*, pp. 507–8 ?

[2] *Brit. J. Psych., Gen. Sect.*, XLII, pp. 192–3.

reply that they have learnt far more from the results of tests and psychological methods of observation than from the study of electro-encephalograms; and that many a delinquent, described by doctors as suffering from 'constitutional defects of cerebral organization,' has been successfully cured by readjusting whatever has caused their 'emotional maladjustments.' In fact it is tempting to remind Mr. Zangwill of the dictum of the most eminent neurologist of our day, Sir Charles Sherrington, who has warned us more than once that "we still know far less about the working of the brain than we do about the working of the mind."

The remaining chapters of the book fall outside the province of the statistical reviewer, and therefore hardly call for comment here. In the main they appear to be excellent. And it may certainly be said that, whether or not Mr. Zangwill's readers agree with his assessments of present-day trends, they will find his account both readable and provocative, and full of interest to those who already have some acquaintance with the work he discusses.

A. F. H.
C. B.

Vocabulaire de la Psychologie. By H. Piéron. Paris: Presses Universitaires de France. Pp. xii + 360. 1,500 *fr.*

Thirty years ago *la Société française de la Psychologie* appointed a committee to prepare a 'vocabulary of psychology.' Unfortunately the project was never actually completed. More recently, however, the Psychological Section of *l'Association de Travailleurs Scientifiques* persuaded several of its members to take up the task afresh. In some ways the *Vocabulaire* resembles Warren's well-known *Dictionary of Psychology*; but (as Professor Piéron points out in his preface) that is already somewhat out of date, and does not meet the French reader's requirements. Moreover, the aim of the French publication is to offer formal definitions chiefly for those terms that are specific to psychology. Hence there are comparatively few definitions of popular or non-technical words.

The selection of concepts to be defined is admirable; and the definitions themselves are as lucid as they are sound. Statistical terminology is dealt with much more thoroughly than in other dictionaries of the same type. But there are one or two trifling slips: for example, the word 'bifactor' is said to be a term used by Burt to designate bipolar factors.

Appended to the dictionary are a number of useful *Annexes*, giving statistical formulæ, the commoner mathematical symbols, abbreviations and symbols used to designate factors, test-results, and the like. The whole publication is admirably produced; and will be useful to British students of psychology as well as to French.

R. F.

Mental Prodigies. By F. Barlow. London: Hutchinson. 1951. Pp. 256. 12*s.* 6*d.*

The author describes this book as "an enquiry into the faculties possessed by persons commonly regarded as 'prodigies' in virtue of their performance in arithmetic, music, memorization, and the like." It deals most fully with the performances of arithmetical prodigies; and a good deal of light is thrown on their methods by the accounts given of the various devices evolved by entertainers who surprise their audiences with their remarkable feats of memory and computation. He contends that the modern psychologist, while insisting on the 'unconscious' nature of many 'emotional' processes, is apt to forget that many 'intellectual' processes may also be carried out 'unconsciously.' He gives numerous instances from his own mental procedures; and reminds us that it was the remarkable performances of Zerah Colburn which first provided Sir William Hamilton with the evidence that led him to develop the theory of unconscious mental processes. The writer himself is well known both in the field of psychical research and as a practical magician; and it is his experiences in these directions that impart an exceptional interest to his discussions. Although the writing is popular and the speculative portions somewhat crude, readers who are interested in the psychology of mathematical thinking will find many of the chapters extremely suggestive.

D. K.

ERRATA

I regret that in the article entitled 'Factor Rotation by the Method of Extended Vectors' in Vol. IV, Part I of this *Journal*, a small computational error occurred in Table II.

The loadings of test 17 should read ·663, ·279, — ·081, — ·213, ·147; that is, the signs in columns III–V are reversed. Line 17 of Table III now becomes ·719, — ·108, ·697, ·592, ·381.

The extended factor loadings for test 17 (Table IV) become 1, ·823, ·247, ·126, ·676, ·362.

This does not alter the positioning of the reference vectors. In most cases it helps to justify their present positions. The values for test 17 in Table V now become ·039, ·292, ·400, ·443, ·217, and row 17 of Table VIII becomes ·325, ·246, ·277, ·391, ·000, ·284, ·108, ·288, ·162, ·597.

J. Sutherland

Vol. V
Part II

The British Journal of Psychology,
Statistical Section

June,
1952

ON THE ADJUSTMENT OF TERMAN-MERRILL I.Q.S TO SECURE COMPARABILITY AT DIFFERENT AGES

By J. A. FRASER ROBERTS and MARGARET A. MELLONE

Burden Mental Research Department, Stoke Park Colony, Bristol,
and Research Department, Royal Eastern Counties Institution, Colchester

I. Introduction. II. *The Data.* III. *The Relation of Variance of I.Q. to C.A.* IV. *The Relation of Mean I.Q. to C.A.* V. *The Use of Cubic Regression Functions for Adjusting I.Q.s to Secure Comparability at Different Ages.* VI. *A Comparison of the Results given by the Two Samples.* VII. *Skewness.* VIII. *Expected Changes in I.Q.* IX. *The Effect of the Adjustments in Removing the Association between I.Q. and C.A.* X. *Summary.*

I. INTRODUCTION

In mental testing, for a wide variety of practical purposes, the Binet Scale occupies a position of special importance. Amongst many advantages is its standardization from the age of 2 years up to adult levels. The items for widely different ages differ in their content ; nevertheless each level merges into the level above ; and this fact, together with all the work that has been devoted to it, enables it to be used as a single scale applicable to all subjects except the very youngest. Yet, even after all these years, the calculation of scores equalized for varying chronological ages is not satisfactory, This is largely due to the use for this purpose of the I.Q. technique.

There is no reason to doubt that the trend of the means of I.Q. is satisfactory. In individuals, or groups of individuals, with an I.Q. close to 100, there is little if any systematic tendency for the figure to rise or fall. I.Q.s differing considerably from 100, however, could only remain constant provided that the variability of mental age increases directly in proportion to chronological age. This is demonstrably far from being true. In introducing their new revision, Terman and Merrill (1937) pointed out that the fluctuations in the standard deviation of I.Q. between different age-groups could not be random, though they were inclined to single out 6 years and 12 years as especially different from the others, with the suggestion that the low variability at 6 years might be due to the nature of the sampling, or to something in the construction of that part of the scale, while the high variability at 12 might have something to do with the onset of puberty.

McNemar (1942) in his statistical volume on the new revision holds that both departures are an artefact, due to the fact that at the 6-year level there are no items of medium difficulty, whereas at 11 and 12 there are too many as compared with other parts of the scale. It is most unlikely that the variations in variance have any particular meaning in terms of real mental growth ; they are doubtless due to the way the scale is constructed and the way the measurements are used. McNemar's suggestion may well afford a satisfactory explanation for the principal features of the fluctuations. In his Appendix B McNemar gives a correction table. The adjustments are large. Thus an obtained I.Q. of 148 is raised to 159 in the C.A. range 4.10 to 6.6 and lowered to 142 in the range 11.6 to 12.5, a difference of 17 points. Similarly an obtained I.Q. of 50 is lowered to 39 in the first age range and raised to 56 in the second. For practical purposes these tables are insufficiently detailed. A child of 4 years 9 months with an obtained I.Q. of 50 is subject to no adjustment. A child with the same I.Q., but one month older, has 11 points deducted. The breaks are too sudden and the adjustments too uniform over a wide range of age.

Goodenough (1942) mentions " spurious age changes of from 15–20 points of I.Q. in the apparent standing of an individual child who ranks from 2–3 S.D. above or below the mean of his group." She obtained standard deviations for yearly age-groups which were very similar to those of the Terman-Merrill sample and suggests that I.Q.s should be adjusted by using the obtained standard deviations at each year level on the basis of a uniform standard deviation of 17. Even this is not as detailed as could be desired. The sampling errors of the obtained figures are relatively large ; and there is also the difficulty of the discontinuity in passing from one year to the next. Clearly, what is desirable is, if possible, some kind of fitted curve, if for no other purpose than smoothing.

It does not follow that all the trouble is due to changing variability of I.Q.s. Some of it, perhaps even a substantial part, might be due to skewness. Skewness which was constant over the whole range might be unfortunate in some ways, but would not affect the comparability of I.Q.s at different ages. There is ample evidence, however, that skewness is in fact very different at different levels. In the Terman-Merrill sample McNemar found that over all ages there was no significant skewness, g_1 being 0·028 ± 0·045. But over the range 6–13 years g_1 was + 0·133 ± 0·061, this being balanced by negative skewness in older and younger subjects. But these are very broad ranges of age. For the magnificent and truly random sample of 1,215 Terman-Merrill I.Q.s collected by Kennedy-Fraser and his colleagues, as described in *The Trend of Scottish Intelligence* (1949), g_1 turns out to be no less than + 0·508 ± 0·070. The range of chronological age was narrow, 95 per cent. of the children falling within the ages 10.5 to 11.7. A reviewer of the Scottish volume in the *Times Educational Supplement* (September 23rd, 1949) said: "Those who have used the Terman-Merrill Scale with children of 11 or over have long suspected that it exaggerates higher mental ages and I.Q.s." He pointed out that while 3 per cent. of I.Q.s in the Scottish sample fell below 70, no less than 10 per cent. exceeded 130. It is only too likely that the pronounced positive skewness at this level contrasts with normality at certain other levels and probably with negative skewness in some of the younger age-groups. This is a particularly troublesome obstacle to securing comparability of I.Q.s at different chronological ages; but, to anticipate our findings, it seems likely that the pronounced positive skewness at about 11 is the chief deviation. This affects particularly very high I.Q.s; and over the bulk of the range of intelligence it would seem that varying skewness is considerably less important than varying variance. This is fortunate, as very large samples would be needed in order to remove fluctuations due to varying skewness—samples so large that they would probably suffice for a new standardization of the scale.

We may describe very briefly the investigation that forced us to undertake the present enquiry; it is typical of many. In connexion with the classification of mental defectives, and in particular with the hypothesis that defectives can be broadly divided into two groups (the evidence was recently summarized by one of us, J. A. F. R., 1950), it was decided to carry out a survey in conjunction with our colleagues at the two Research Departments. In four areas children of certain ages having Terman-Merrill I.Q.s ranging from 35 to 60 were ascertained as completely as possible. Part of the plan was to test on the same scale all their brothers and sisters of school age. At once the difficulty of the lack of independence of C.A. and I.Q. presented itself. Subjects of I.Q. 35–60 were considerably more numerous at 12 than at 8. The I.Q.s of the brothers and sisters should have been independent of C.A., but they were not. The mean I.Q. was well below 100; and so, as on the whole variability of I.Q. increases with age, there was a highly significant fall of three-quarters of a point per year. A contingency table based on our sample of sibs and cousins, as shown in Table I, revealed a gross departure from independence of I.Q. and C.A., χ^2 for 28 degrees of freedom being 76·4: (this is mentioned again in a later section). Hence it seemed essential to devise suitable adjustments for the I.Q.s of subjects and sibs. We also hoped that, if we could make some contribution to this problem, the results would be of use to those who are interested in children whose intelligence departs widely from the mean; for in the school medical service or at child guidance clinics it is generally recognized that some kind of age adjustment of I.Q.s must be made.

II. THE DATA

Our own sample consists of the sibs of the mental defectives mentioned in the preceding section, together with a further sample of first cousins of the same subjects, which was obtained from two of the three areas based on Colchester. Sampling was complete in that practically all sibs of school age were ascertained and tested on Form L. We made no attempt to go outside these limits; before school age, and more particularly after the compulsory minimum leaving age of 15, it is almost impossible to make the testing complete. The range is thus 5 years 0 months to 14 years 11 months. In view of the nature of the sample there were some very low I.Q.s. We decided to omit from the analysis 4 that fell below 31, leaving a total number available of 689.

The sample appears to be entirely suitable for investigating the relation of I.Q. and C.A. With the single restriction that children came into the sample because they were sibs (or cousins) of subjects with I.Q.s ranging from 35 to 60, ascertainment was complete. It seems inconceivable that this method of selection could introduce any artificial association with C.A. The mean I.Q. is naturally lower than that of the general population, but that is

automatically allowed for in the analysis. There is no reason to suppose that the variances would be abnormal ; or what is really important, the ratio of the variance at one age to that at another. The method of selection introduces, of course, negative skewness, but we had little hope of doing anything about this ; rather we hoped that, having corrected for variation of variance, the remaining association due to skewness would prove relatively unimportant.

Table I shows C.A.s in one-year groups and I.Q.s in groups of 3 points. All calculations have been made in these units.

TABLE I. AGES AND I.Q.S : OWN SAMPLE

Sibs and Cousins of School Age of a Group of Mental Defectives of I.Q. Range 35–60, *showing C.A. and Terman-Merrill I.Q.*

I.Q.*	Chronological Age										
	5.0–5.11	6.0–6.11	7.0–7.11	8.0–8.11	9.0–9.11	10.0–10.11	11.0–11.11	12.0–12.11	13.0–13.11	14.0–14.11	
31–33				1				1			2
34–36				1						1	2
37–39						1	1				2
43–45						2	1	1			4
46–48			1					3	2		6
49–51	1						1		1	1	4
52–54	1	1	1	1		1	2		3	1	11
55–57					1	1		4			6
58–60	1		1		2		3	2	1	3	13
61–63	1					2	1	1	2	3	10
64–66	2	2		1	1	3	4		1	2	16
67–69		2	1	1	1	3	5	3	3	3	22
70–72	2	2	1	4	9	4	3	2	2	1	30
73–75	2	1	7	5			3	3	2	2	25
76–78	1	3	2	5	3	4	2	2	5	6	33
79–81	3	3	2	2	4	4	5	5	3	3	34
82–84	1	3	9	9	5	8	6	4	5	4	54
85–87	4	3	8	5	7	4	10	5	6	3	55
88–90	2	7	1	7	3	4	3	5	5	7	44
91–93	3	8	3	4	5	3	4	3	7	5	45
94–96	2	6	4	4	1	6	6	6	9	10	54
97–99	6	6	2	3	5	5	3	6	2	4	42
100–102	10	8	2	3	6	3	1	4	3	2	42
103–105	2	6	2	4	4		3		4	3	28
106–108	4	3	2	4	7	3	4	3	2		32
109–111	3	3	2	1	1	4	2	3	1		20
112–114	1								4	1	6
115–117	2	2	2		3	5			1	1	16
118–120			1	1		1	2	1	1		7
121–123			1			1	1			1	4
124–126					1	1	1				3
127–129		1			1					1	3
130–132			1		1	2	2				6
133–135				1			1				2
136–138				1	1	1					3
142–144						1					1
154–156									1		1
163-165							1				1
Total	54	70	56	68	72	77	81	67	76	68	689
Mean I.Q.	91·50	92·47	88·46	87·37	91·33	89·90	86·78	82·96	87·18	84·76	

* Blank rows not shown.

We have carried out duplicate calculations on the Terman-Merrill sample (Form L), using for this purpose the 2,000 children in the age range 5–14. The figures used are shown in Table II, taken from Tables 5 and 7 in *Measuring Intelligence*. Those who feel doubtful about our sample because the children are sibs of defectives will be reassured by the remarkably close concordance in essentials of the results for two such very different samples.

TABLE II. MEANS AND VARIANCES: TERMAN AND MERRILL'S SAMPLE

Age	Number	Mean	Variance
5	200	102·7	202·14*
6	203	101·0	156·25
7	202	101·8	262·44
8	203	102·4	249·64
9	204	104·3	268·96
10	201	104·5	272·25
11	204	104·3	324·00
12	202	103·6	400·00
13	204	104·0	320·41
14	202	100·9	259·21
Total	2,025		

* Based on 219 children.

III. THE RELATION OF VARIANCE OF I.Q. TO C.A.

(*A*) *Our Own Sample.* Using the data of Table I, it turns out that the relation of variance of I.Q. to C.A. is described by a cubic regression function. The equation is :

$$Y = 31{\cdot}3929 - 8{\cdot}61052x + 3{\cdot}251528x^2 - 0{\cdot}24112747x^3,$$

where Y is the variance of I.Q. in units of grouping of 3 points of I.Q. and x is age in yearly age-groups ; 5.0–5.11 = 1, 6.0–6.11 = 2, etc.

In calculating the product terms the variance in each array is multiplied by the number in the array and not by one less. The mean variance is obtained by dividing the sum of these weighted terms by the total number in the sample. In all other calculations also the weights are n_p, the number in the array, and not $n_p - 1$.

Table III shows (in units of grouping) the observed variances and the expected variances, using respectively the linear, quadratic, and cubic functions.

The total weighted sum of squares of the variances, in the above units, is 64,795·3, of which the cubic regression function accounts for 48,290·0. The problem of examining the significance of the successive terms of the polynomial and of the remainder is a somewhat unusual one, as regressions of variances on other variables are seldom used (though they would seem to have useful applications in connexion with psychological measurements). We owe the solution to Professor R. A. Fisher, who gave it to one of us fifteen years ago in connexion with a similar problem (Roberts, Norman, and Griffiths, 1935). A somewhat amplified account is now presented.

The mean square, derived by dividing the weighted sum of squares of deviations of observed variances from the regression line by $k - r$, where k is the number of arrays and $r - 1$ the order of the fitted polynomial, is :

$$\Sigma n_p(y_p - Y_p)^2/(k - r). \qquad (1)$$

The theoretical sampling variance of an estimate y of a true variance Y, based on a sample of n, is $2Y^2/n$. The mean square (1) therefore estimates approximately the mean value of $2Y^2$, taken over the k arrays, namely, $2\Sigma Y_p^2/k$. The approximation arises from the fact that the observed variances were originally weighted only according to the number of observations and not according to estimates of their true values. Any discrepancy due to this cause, however, would be very small.

TABLE III. OBSERVED AND EXPECTED VARIANCES IN ARRAYS

Unit of Grouping, 3 points of I.Q.

Age	Age-Group	No.	Observed Variance	Expected Variance		
				Linear Function	Quadratic Function	Cubic Function
5.0–5.11	1	54	29·2358	28·9285	18·8888	25·7928
6.0–6.11	2	70	20·4822	30·6454	26·8724	25·2490
7.0–7.11	3	56	30·1857	32·3623	33·3145	28·3147
8.0–8.11	4	68	34·3115	34·0792	38·2152	33·5432
9.0–9.11	5	72	33·5837	35·7961	41·5743	39·4876
10.0–10.11	6	77	51·8964	37·5130	43·3921	44·7013
11.0–11.11	7	81	52·4694	39·2299	43·6683	47·7375
12.0–12.11	8	67	38·7725	40·9468	42·4031	47·1493
13.0–13.11	9	76	39·2554	42·6637	39·5964	41·4901
14.0–14.11	10	68	32·3951	44·3806	35·2482	29·3131
Total		689				

Mean Variance 36·987204

Table IV shows the analysis of variance of variance between arrays. In examining the significance of the linear term the mean square is compared with twice the square of the mean variance, the null hypothesis being that there is no significant departure from the horizontal line. The extra mean square due to fitting the quadratic is compared with 2/10 times the sum of the expected variances given by the linear function, i.e., the fifth column in Table III. Similarly, for the cubic term the expected variances on the quadratic function are used, and for the remainder, after fitting the cubic, the expected variances on the cubic function.

TABLE IV. REGRESSION OF VARIANCE OF I.Q. ON C.A.

Analysis of Variance of Variance between Arrays : Own Sample

Variation of Variance of I.Q. between arrays due to :	Degrees of Freedom	Sum of Squares	Mean Square	Corresponding Theoretical Variance (*D.* of *F.* = ∞)	Variance Ratio	Significance
Linear regression	1	15,842	15,842	2,736	5·79	+
Addition of quadratic term	1	20,843	20,843	2,736	7·62	++
Addition of cubic term ..	1	11,605	11,605	2,755	4·21	+
Remainder	6	16,505	2,751	2,774	0·99	−
Total	9	64,795				

It will be seen that there is a significant improvement in the fit at each step up to the cubic and that the remainder after fitting the cubic is non-significant and is in fact almost exactly equal to expectation. The relation of variance of I.Q. to C.A. in our sample is, therefore, summarized with high efficiency by a cubic regression function. The observed means of the arrays and the fitted curve are shown in Fig. 1.

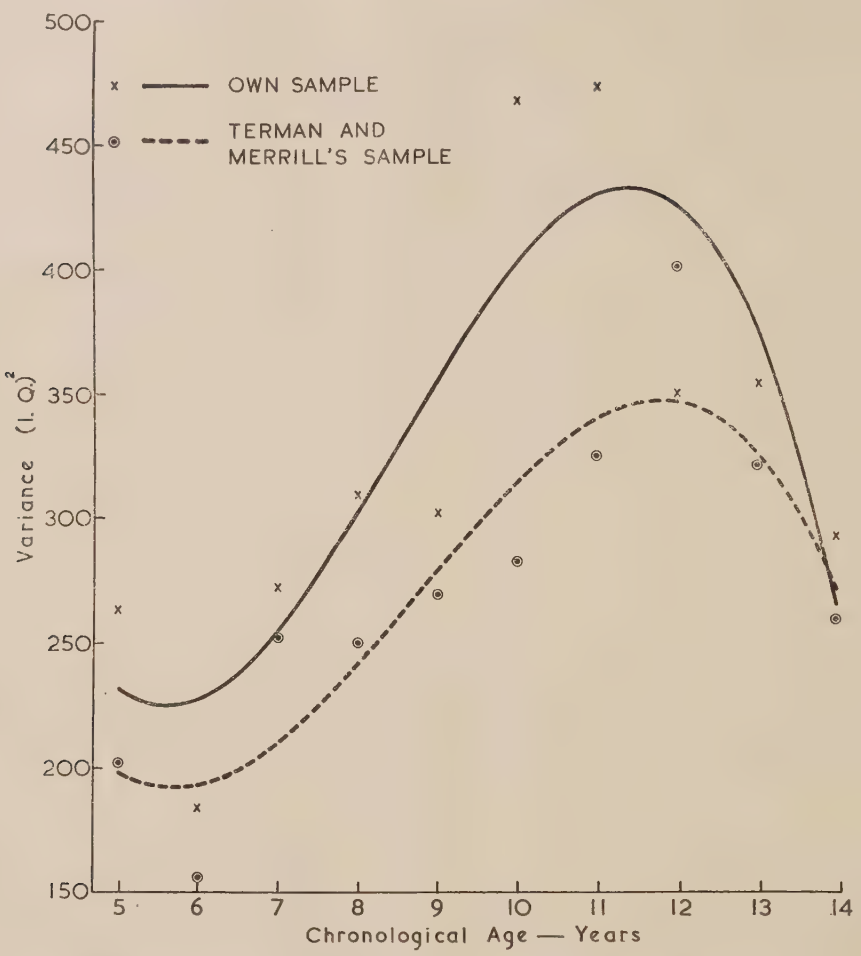


FIG. 1.—Regression of variance of I.Q. on C.A. Means of yearly arrays, with fitted cubic functions.

(*B*) *Terman-Merrill Sample.* The procedure of the previous section was repeated with the Terman-Merrill sample, using the figures given in Table II, which is taken from Tables 5 and 7 of *Measuring Intelligence*. A slight adjustment was necessary at age 5, as the raw mean was based on 200 children, whereas the standard deviation was based on the larger number of 219.

Once again the regression of variance of I.Q. on C.A. is described by a cubic function. The equation is :

$$Y = 234{\cdot}410 - 54{\cdot}3051x + 19{\cdot}49819x^2 - 1{\cdot}3715113x^3,$$

where Y is variance of I.Q. and x age in years, 5.0–5.11 = 1, etc.

The steps in the fitting are shown in Table V, which corresponds to Table IV above.

TABLE V. REGRESSION OF VARIANCE OF I.Q. ON C.A.
Analysis of Variance of Variance : Terman-Merrill Sample

Variation of Variance of I.Q. between arrays due to :	Degrees of Freedom	Sum of Squares	Mean Square	Corresponding Theoretical Variance ($D.$ of $F. = \infty$)	Variance Ratio	Significance
Linear regression	1	4,113,096	4,113,096	147,457	27·89	++
Addition of quadratic term	1	1,058,490	1,058,490	151,482	6·99	++
Addition of cubic term ..	1	1,187,598	1,187,598	163,917	7·25	++
Remainder	6	1,913,304	318,884	153,679	2·08	—
Total	9	8,272,488				

There is a highly significant improvement at each step up to the cubic. The remainder, after fitting the cubic, is almost twice expectation—just below the level of significance. The numbers in the year groups were so nearly equal that the fitting could be carried out by orthogonal polynomials. Hence it was easy to see whether additional terms might not secure some further improvement. Actually, the biquadratic and quintic terms removed less than one-sixth and one-fifth respectively of the residual sum of squares. Thus, as in our own sample, the regression of variance of I.Q. on C.A. is described with high efficiency by a cubic function. The observed means of arrays and the fitted curve are shown in Fig. 1.

IV. THE RELATION OF MEAN I.Q. TO C.A.

(*A*) *Our Own Sample.* From Table I we find : Mean I.Q., 88·2206 ; regression of I.Q. on C.A., — 0·785849 point per year. The fitting of the regression is shown in Table VI.

TABLE VI. REGRESSION OF I.Q. ON C.A.
Analysis of Variance : Units of I.Q. Own Sample

Variation of I.Q. due to :	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio
Linear regression	1	3,319	3,319	9·97
Deviations from linear regression	8	2,413	302	0·91
Within years	679	226,094	333	
Total	688	231,826		

The fall of mean I.Q. with advancing age is highly significant. It is wholly accounted for by linear regression, the remainder between years being smaller than the mean square within years. The fall is clearly a reflection of the general tendency over the whole range for variance to increase with age. If, as must be assumed, the fall is due to the changing variance, as revealed in the previous sections, it might be asked why the changing mean of the sample should not also show a cubic relationship to C.A. Actually the array means, shown at the bottom of Table I, do suggest this. The mean rises to a maximum at age 6, falls to a minimum at about 11 or 12, and then rises again ; this is just the reciprocal relation that would be expected. But it will also be seen that there is much irregularity ; and with numbers like these and with such small expected changes in the mean it is not surprising to find that a straight line accounts for the whole of the significant association.

TABLE VIII. CORRECTION TABLE FOR TERM

Chronological Age Years	Months	Factor	Factor* Own Sample	Subtract −9	−8	−7	−6	−5	−4	−3	−2
5	0–2	1·0423	1·0666								50–6
	3–5	1·0600	1·0875							50–61	62–7
	6–8	1·0731	1·1025						50–55	56–68	69–8
	9–11	1·0813	1·1114						50–59	60–72	73–8
6	0–2	1·0849	1·1141						50–61	62–73	74–8
	3–5	1·0840	1·1110						50–61	62–73	74–8
	6–8	1·0789	1·1028						50–58	59–71	72–8
	9–11	1·0703	1·0903						50–53	54–67	68–8
7	0–2	1·0586	1·0743							50–60	61–7
	3–5	1·0446	1·0558								50–6
	6–8	1·0288	1·0356								50
	9–11	1·0118	1·0144								
8	0–2	·9940	·9928								
	3–5	·9760	·9714								
	6–8	·9580	·9506								139–1
	9–11	·9405	·9305							146–150	129–1
9	0–2	·9235	·9116						149–150	136–148	123–1
	3–5	·9074	·8939						141–150	130–140	120–1
	6–8	·8921	·8776					145–150	136–144	127–135	117–1
	9–11	·8779	·8628				149–150	140–148	132–139	124–131	116–1
10	0–2	·8648	·8494				144–150	137–143	129–136	122–128	115–1
	3–5	·8530	·8377			148–150	141–147	134–140	127–133	121–126	114–1
	6–8	·8423	·8275			145–150	138–144	132–137	126–131	119–125	113–1
	9–11	·8330	·8190		148–150	142–147	136–141	130–135	124–129	118–123	112–1
11	0–2	·8250	·8122		146–150	141–145	135–140	129–134	123–128	118–122	112–1
	3–5	·8184	·8072	150	145–149	139–144	134–138	128–133	123–127	117–122	112–1
	6–8	·8133	·8042	149–150	144–148	138–143	133–137	128–132	122–127	117–121	112–1
	9–11	·8097	·8031	148–150	143–147	138–142	132–137	127–131	122–126	117–121	111–1
12	0–2	·8077	·8042	148–150	143–147	137–142	132–136	127–131	122–126	117–121	111–1
	3–5	·8075	·8077	148–150	142–147	137–141	132–136	127–131	122–126	116–121	111–1
	6–8	·8092	·8139	148–150	143–147	138–142	132–137	127–131	122–126	117–121	111–1
	9–11	·8129	·8232	149–150	144–148	138–143	133–137	128–132	122–127	117–121	112–1
13	0–2	·8190	·8361	150	145–149	139–144	134–138	128–133	123–127	117–122	112–1
	3–5	·8277	·8534		147–150	141–146	135–140	130–134	124–129	118–123	112–1
	6–8	·8394	·8760		150	144–149	138–143	132–137	125–131	119–124	113–1
	9–11	·8548	·9056			148–150	141–147	134–140	128–133	121–127	114–1
14	0–2	·8746	·9442				147–150	139–146	131–138	123–130	115–1
	3–5	·8998	·9957					148–150	138–147	128–137	118–1
	6–8	·9320	1·0659							140–150	126–1
	9–11	·9736	1·1662								

*For comparison.

For I.Q.s between 50 and 150 the correction is read off directly. For example, for I.Q. 135 at 11 years 10 months subtract 6 points ; for I.Q. 70 at 9 years 1 month add 3 points.

RILL I.Q.S (BASED ON TERMAN-MERRILL SAMPLE)

		Add									
1	0	1	2	3	4	5	6	7	8	9	10
91	92–114	115–138	139–150								
94	95–111	112–128	129–144	145–150							
96	97–109	110–123	124–137	138–150							
96	97–109	110–121	122–133	134–146	147–150						
97	98–108	109–120	121–132	133–144	145–150						
97	98–108	109–120	121–132	133–144	145–150						
96	97–109	110–122	123–134	135–147	148–150						
95	96–110	111–124	125–138	139–150							
94	95–111	112–128	129–145	146–150							
91	92–114	115–136	137–150								
85	86–120	121–150									
80	61–145	146–150									
	50–150										
50	83–123	50–82									
38	92–114	68–91	50–67								
28	95–111	78–94	61–77	50–60							
22	97–109	84–96	71–83	58–70	50–57						
19	98–108	87–97	77–86	66–76	55–65	50–54					
16	99–107	90–98	80–89	71–79	62–70	53–61	50–52				
15	99–107	91–98	83–90	75–82	67–74	58–66	50–57				
14	100–106	92–99	85–91	78–84	70–77	63–69	55–62	50–54			
13	100–106	93–99	86–92	80–85	73–79	66–72	59–65	52–58	50–51		
12	100–106	94–99	88–93	81–87	75–80	69–74	62–68	56–61	50–55		
11	101–105	95–100	89–94	83–88	77–82	71–76	65–70	59–64	53–58	50–52	
11	101–105	95–100	89–94	83–88	78–82	72–77	66–71	61–65	55–60	50–54	
11	101–105	95–100	90–94	84–89	79–83	73–78	68–72	62–67	57–61	51–56	50
11	101–105	95–100	90–94	85–89	79–84	74–78	69–73	63–68	58–62	53–57	50–52
10	101–105	96–100	90–95	85–89	80–84	75–79	69–74	64–68	59–63	54–58	50–53
10	101–105	96–100	90–95	85–89	80–84	75–79	70–74	64–69	59–63	54–58	50–53
10	101–105	96–100	91–95	85–90	80–84	75–79	70–74	65–69	59–64	54–58	50–53
10	101–105	96–100	90–95	85–89	80–84	75–79	69–74	64–68	59–63	54–58	50–53
11	101–105	95–100	90–94	85–89	79–84	74–78	69–73	63–68	58–62	53–57	50–52
11	101–105	95–100	90–94	84–89	79–83	73–78	68–72	62–67	57–61	51–56	50
11	101–105	95–100	89–94	83–88	77–82	72–76	66–71	60–65	54–59	50–53	
12	100–106	94–99	88–93	82–87	75–81	69–74	63–68	57–62	51–56	50	
13	100–106	93–99	86–92	79–85	73–78	66–72	59–65	52–58	50–51		
14	100–106	92–99	84–91	76–83	68–75	60–67	52–59	50–51			
17	99–107	89–98	79–88	69–78	59–68	50–58					
25	96–110	81–95	67–80	52–66	50–51						
50	85–121	50–84									

or I.Q.s outside the range 50–150 multiply deviation from 103 by factor shown in column 2 and add to or btract from 103.

or example, to adjust I.Q. 173 at 12 years 2 months : $103 + (70 \times 0{\cdot}8077) = 160$. For I.Q. 45 at 10 ars 1 month : $103 - (58 \times 0{\cdot}8648) = 53$.

(*B*) *Terman-Merrill Sample.* The data of Table II show that there is no significant heterogeneity in the yearly raw means : Mean I.Q., 102·9526 ; regression of I.Q. on C.A., + 0·11996 point per year. The fitting of the regression line is shown in Table VII. The

TABLE VII. REGRESSION OF I.Q. ON C.A.

Analysis of Variance : Terman-Merrill Sample

Variation of I.Q. due to :	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio
Linear regression	1	240	240	0·88
Deviations from linear regression	8	3,258	407	1·50
Within years	2,015	546,762	271	
Total	2,024	550,260		

mean square due to linear regression is smaller than the mean square within arrays, and the residual mean square between arrays not significantly greater. Hence the mean I.Q. at all ages is adequately represented by the single general mean given above.

V. THE USE OF CUBIC REGRESSION FUNCTIONS FOR ADJUSTING I.Q.S TO SECURE COMPARABILITY AT DIFFERENT AGES

The adjustments based on the Terman-Merrill sample will be described first ; they are very simple. It was thought that it would be sufficient to provide corrections at intervals of three months of C.A. Hence, using the cubic regression function, an expected variance was calculated for 5 years 1 month, 5 years 4 months, and so on up to 14 years 10 months. From these the standard deviations were obtained. The standard age was fixed at 8 years. The expected standard deviation at 8 years (15·024) was divided successively by the 40 expected standard deviations at the three-month intervals, so giving column 2 of Table VIII. The mean I.Q. of the Terman-Merrill sample is, rather conveniently, 103·0 to one place of decimals, hence to adjust any I.Q. all that is necessary is to multiply the deviation from 103 by the appropriate factor of column 2, the answer being added to or subtracted from 103. Table VIII enables the correction to be read off immediately, given C.A. and I.Q. : thus, for an I.Q. of 135 at 11 years 10 months, 6 points of I.Q. are subtracted. There seems to be general agreement that mean I.Q. on the Terman-Merrill Scale is in fact slightly greater than 100 ; in the Scottish Survey sample, for instance, it was 102·52. Hence our use of 103 for the mean I.Q. instead of 100.

For our own sample a series of factors at three-month intervals was calculated in the same way. It is possible to use these factors in two ways. First, deviations can be measured from the same population mean of 103 as was used with the Terman-Merrill sample. An alternative plan is to make no assumption as to the population mean and to make the corrections from the fluctuating mean of the sample itself. This involves using the regression of mean I.Q. on C.A. given in a preceding section. At each age the deviation is measured from the expected mean at that age. In addition there is added to or subtracted from each I.Q. the amount by which the expected mean differs from that at the standard age of 8 years.

This second plan leads to a result of some interest, as it is possible to estimate, simply from our sample of low I.Q.s, what the mean for the general population should be. It turns out that the I.Q. which most nearly corresponds to zero adjustment is 101 and there is little departure from this figure except in the oldest age-groups, 13–14 years, and to a lesser extent in the very youngest children. This result, also, seems to indicate that our sample is a good one.

The factors derived from our own sample are given in the third column of Table VIII.

VI. A COMPARISON OF THE RESULTS GIVEN BY THE TWO SAMPLES

Our own sample shows considerably more variability of I.Q. than does the Terman-Merrill. We are inclined to believe that our larger variances are probably nearer those of the general population. One indication of this is given by a comparison with the Scottish 1,215. The mean age of the Scottish children was 10 years 11·75 months (with a narrow spread). The observed standard deviation was 20·1 ; and the expected figure on the basis of the cubic function derived from our sample is 20·7. For the purpose of age-correction, however, the absolute difference is immaterial. As already pointed out it is only the ratio of the variance at one age to that at another that matters. Actually the corresponding factors of columns 2 and 3 of Table VIII are very similar.

Comparing the corrections which adjust I.Q.s to age 8, these are very much the same from 5 years to 13 years 7 months and for the full range of possible I.Q.s. For example, taking the Terman-Merrill factors and our own factors, given a population mean of 103, the corrections over the range 50 I.Q. to 150 I.Q. vary between + 10 and − 9 points. There are 2 × 35 comparisons for these two levels of I.Q. at the three-monthly intervals of C.A. from 5.1 to 13.7 inclusive. The differences in the correction turn out as follows :

0 point,	25	times
1 ,,	37	,,
2 points,	8	,,
	70	

Thus, over a very wide range of I.Q.s and this wide span of age, the difference in the correction is nearly always 0 or 1 point, occasionally 2 points, and never more.

From 13 years 7 months upwards there is an increasing discrepancy due to the steeper fall of the curve derived from our sample. Here the Terman-Merrill sample seems to provide the better guide. Corrections based on our sample lead, in the oldest children, to impossible results ; in children close to 15 years an I.Q. 50 points from the mean would appear to be changing at the rate of 20 points a year, which is absurd. Doubtless we happened by chance to have in our relatively small sample a particularly non-variable set of oldest children.

The alternative plan, mentioned in the previous section, of using the fluctuating mean of the sample itself instead of the constant figure of 103 yields a set of corrections which diverge slightly more from the Terman-Merrill in the youngest children, but otherwise for I.Q. 50 correspond closely up to the extreme age of 14 years 10 months. This closer correspondence is secured, however, at the expense of a greater discrepancy at I.Q. 150 from 13.0 years upwards.

Table IX shows the corrections for I.Q. 50 and I.Q. 150 at yearly intervals from 5 years 7 months to 13 years 7 months on the basis of the Terman-Merrill sample and the two ways of using the figures from our own sample.

TABLE IX. COMPARISON OF CORRECTIONS

C.A.	I.Q. 50			I.Q. 150		
	Terman-Merrill Sample	Own Sample		Terman-Merrill Sample	Own Sample	
		Mean 103	Fluctuating Sample Mean		Mean 103	Fluctuating Sample Mean
5.7	−4	−5	−6	+3	+5	+4
6.7	−4	−5	−5	+4	+5	+5
7.7	−2	−2	−2	+1	+2	+2
8.7	+2	+3	+2	−2	−2	−2
9.7	+6	+6	+6	−5	−6	−6
10.7	+8	+9	+9	−7	−8	−9
11.7	+10	+10	+10	−9	−9	−9
12.7	+10	+10	+10	−9	−9	−8
13.7	+9	+7	+9	−8	−6	−4

It seems sufficient to use the Terman-Merrill sample corrections ; and this is what we intend to do in the survey mentioned in the introduction. But the correspondence is so close that except in the oldest children it hardly matters which is used. The close correspondence is very satisfactory, and gives some confidence in the method employed.

VII. SKEWNESS

The problem of skewness has already been discussed in the introduction. Our sample of 689 children is too small for any firm conclusions ; but the results shown in Table X are suggestive.

The method of selection of these children must inevitably have introduced considerable negative skewness. It seems probable that this has partially neutralized the expected positive skewness in the age-group 10–12½. (For the Scottish Survey children g_1 was 0·508 ± 0·070.) The youngest age-group yields the only significant value ; and it would seem that at this level the skewness in the general population is very unlikely to be positive. We have one additional figure for a small sample of 120 6-year children which has recently been studied by one of us (M. A. M.). This is a random sample, chosen, as in the Scottish sample, by day of birth in the month, and yields $g_1 = -0{\cdot}171 \pm 0{\cdot}221$.

TABLE X. SKEWNESS

Own Sample of 689 Children

Age	Number	g_1
5.0–7.5	149	− 0·549 ± 0·199
7.6–9.11	171	+ 0·075 ± 0·186
10.0–12.5	201	+ 0·213 ± 0·172
12.6–14.11	168	− 0·021 ± 0·187
Total	689	− 0·013 ± 0·093

These results are in broad agreement with those of McNemar ; and it would seem that the chief effect is the pronounced positive skewness at about 10 to 12 years, this being diluted by a more normal distribution in younger and older children, together with, it may be, negative skewness in the youngest.

Our sample of I.Q.s is well below the mean and it may be that the corrections based on the varying variance would fail to adjust adequately the I.Q.s of clever children at ages, say, 10 to 12, but for samples of somewhat dull subjects, or even for the general population, the corrections elaborated in previous sections may be expected to remove the bulk of the undesired association of I.Q. with C.A. This is treated in a later section.

Incidentally, significant kurtosis does not appear to be a feature of the distribution of Terman-Merrill I.Q.s. To quote the Scottish Survey figures once again, g_2 is not, in this large sample, significantly different from zero.

VIII. EXPECTED CHANGES IN I.Q.

It is of some interest to follow the expected changes in I.Q. at various levels. A selection is shown in Fig. 2. Needless to say, anything like a smooth curve could only be expected on the basis of large samples. The adjustments of Fig. 2 are based on the Terman-Merrill sample. It will be seen that on the whole the expected changes are not unreasonable, even for I.Q.s far from the mean. Thus an I.Q. of 52 at 5 years 1 month may be expected to rise to a maximum of 54 at 5.7 ; remain constant for a year ; fall by about 2 points a year to age 8 ; thereafter fall steadily at the rate of 4 points a year to 11.1 ; remain constant to 13.1 ; and finally rise to 49 again. From 14.7 or so the rise becomes suspiciously great. Whether in fact these actual changes will be shown (on large samples) can only be determined by repeated measurements of the same individuals.

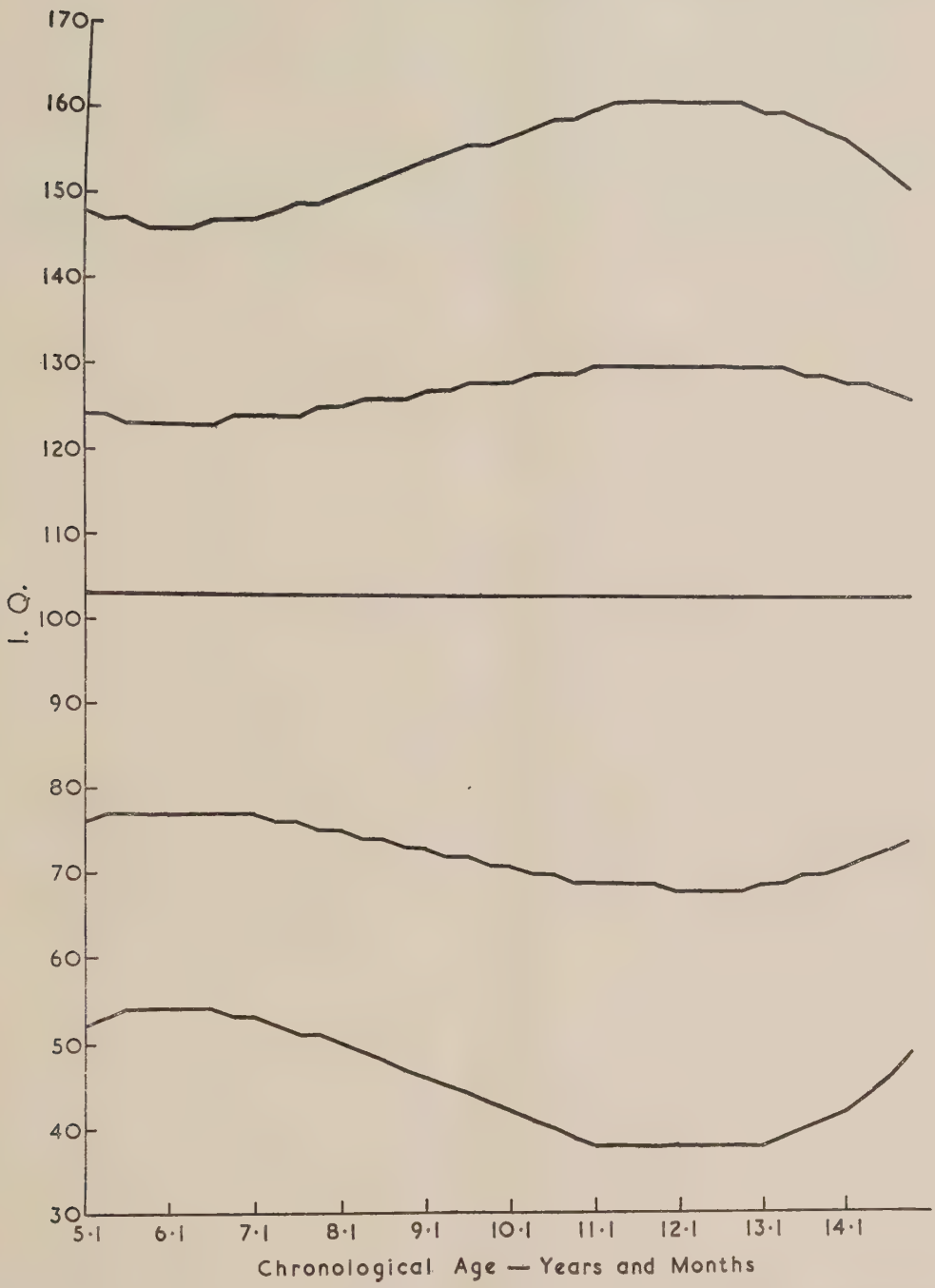


Fig. 2.—Changes expected in I.Q. with changes in age. Five selected levels. Based on adjustments calculated from Terman and Merrill's sample.

IX. THE EFFECT OF THE ADJUSTMENTS IN REMOVING THE ASSOCIATION BETWEEN I.Q. AND C.A.

A contingency table can be formed from the data of Table I for our own sample of 689. Two-year age-groups have been employed. Using these for the I.Q.s of Table I and the three sets of corrected I.Q.s obtained by using the three methods, it was found that in order to secure minimum expectations of 5 in all cells I.Q.s of 60 and under had to be grouped together, as well as I.Q.s of 115 and above. This left in between them 18 of the 3-point I.Q. groups of Table I. It seemed reasonable to add these in threes, so all contingency tables are 8×5, giving 28 degrees of freedom. The results of the tests for independence are shown in Table XI.

TABLE XI. CONTINGENCY TABLES

Tests of Independence of C.A. and I.Q. for Obtained I.Q.s and Adjusted I.Q.s

Variate	χ^2
Obtained I.Q.s	76·36
Adjusted I.Q.s—	
1. Terman-Merrill sample	48·79
2. Own sample : mean I.Q. 103..	50·95
3. Own sample : fluctuating sample mean ..	37·46

For 28 degrees of freedom : 5 per cent. point, 41·34 ; 1 per cent. point, 48·28.

The unadjusted I.Q.s show a marked and highly significant departure from independence of C.A. The adjustment, using either the Terman-Merrill sample figures or our own sample figures, and assuming the same constant population mean of 103, effects a great improvement, χ^2 being reduced to just about the 1 per cent. level of significance. The figures based on our own sample, using the fluctuating mean of the sample itself, give an even better result, χ^2 being now below the 5 per cent. level of significance. The last result probably means that the second method of calculation also removes some feature due to a peculiarity of the sample. That adjustments based on a totally different sample, the Terman-Merrill, should produce so great a measure of improvement is, however, even more satisfactory. It is true that the ends of the distribution have to be rather coarsely grouped, especially in the upward direction, but it does appear that over most of the range of I.Q.s a simple adjustment based solely on fitting a curve of variance of I.Q. on C.A. does in fact eliminate the great bulk of the troublesome lack of independence of C.A. Variable skewness, it would appear, is much less important.

X. SUMMARY

1. Terman-Merrill I.Q.s, except near the mean, are far from being independent of chronological age. This might be due to variations in variance at different C.A.s, as has often been pointed out. It might also be partially due to variations in skewness. Kurtosis does not appear to be variable or abnormal.

2. This paper is based on the results (Form L) of a complete sample of 689 sibs and cousins of school age of a sample of mental defectives of I.Q. range 35–60. Corresponding calculations have been carried out on the figures given for the sample on which the present revision was based, as given in *Measuring Intelligence*. The range of C.A. is 5 years 0 months to 14 years 11 months.

3. The regression of variance of I.Q. on C.A. is fitted in both samples by a cubic regression function, the deviations from which are well below the level of significance. Except in the oldest children the curves are reasonably parallel.

4. The cubic regression functions are used for obtaining a series of factors by which the I.Q. at given C.A. can be adjusted to correspond to I.Q. at 8 years. These have been calculated at 3-month intervals from 5 years 1 month to 14 years 10 months. Table VIII enables the adjustment in points of I.Q. to be read off directly given obtained I.Q. and C.A.

5. The adjustments obtained by using the two samples are closely similar, except in the oldest children. Here the discrepancy is probably due to a peculiarity in our sample and it is concluded that adjustments based on the original Terman-Merrill sample should be sufficient for all practical purposes.

6. Our own sample is too small for the adequate study of variations in skewness, but it is likely that pronounced positive skewness at about 10 to 12 years is diluted by reasonable normality in older and younger children ; and there may be real negative skewness in the youngest children. Except in samples of clever children in the age range mentioned, adjustment for varying variance appears to leave relatively little remaining important lack of independence of I.Q. and C.A.

7. Contingency tables based on our sample show that the gross lack of independence of I.Q. and C.A. is reduced by the adjustments almost to the level of non-significance, even when these adjustments are based on the quite different Terman-Merrill sample.

Acknowledgements. Our own sample consists of I.Q.s of children tested in connexion with a survey the results of which it is hoped to publish shortly in conjunction with our colleagues. In addition to the tests carried out by one of us (M. A. M.) we are indebted for many results to Mrs. M. D. Jenner, Mrs. A. E. Sedgley, and Miss E. H. Davidson. We are also indebted to Dr. P. Armitage, of the Medical Research Council's Statistical Research Unit, London School of Hygiene and Tropical Medicine, for much helpful advice. Our thanks to Professor R. A. Fisher for the method of testing the significance of deviations from the fitted regression lines have been expressed in the text. The figures were drawn by Mrs. M. G. Young.

REFERENCES

1. Goodenough, F. L. (1942). ' Studies of the 1937 revision of the Stanford-Binet Scale ; variability of the I.Q. at successive age-levels.' *J. Educ. Psych.*, XXXIII, 241–51.
2. McNemar, Q. (1942). *The Revision of the Stanford-Binet Scale.* Boston : Houghton Mifflin.
3. Roberts, J. A. F., Norman, R. M., and Griffiths, R. (1935). ' Studies on a child population: Definition of the sample, method of ascertainment, and analysis of the results of a group intelligence test.' *Ann. Eugen.*, VI, 319–38.
4. Roberts, J. A. F. (1950). ' The genetics of oligophrenia.' *Proc. Congrès Inter. de Psychiatrie*, VI. Paris : Hermann.
5. Scottish Council for Research in Education (1949). *The Trend of Scottish Intelligence.* London : University of London Press.
6. Terman, L. S., and Merrill, M. A. (1937). *Measuring Intelligence.* London : Harrap.

Vol. V
Part II

The British Journal of Psychology,
Statistical Section

June,
1952

AN EMPIRICAL STUDY OF POLITICAL, RELIGIOUS, AND SOCIAL ATTITUDES

By M. SANAI

Department of Psychology, University College, London

I. *Problem.* II. *Methods.* III. *Results.* IV. *Group Differences in the General Factor.* V. *Distribution of Degrees of Certainty.* VI. *Summary.*

I. PROBLEM

Generality and Specificity. In the study of social attitudes one of the most fundamental issues is their degree of generality or specificity. Many writers appear to hold that attitudes are always narrow in their range, each representing a highly specific tendency to make a particular response to a particular situation ; others, however, believe that certain attitudes arise from emotional dispositions of a fairly general type, and are thus liable to extend over wide fields of opinion or behaviour. If, in spite of long-drawn-out discussions, no decisive conclusion has been reached, that may perhaps be attributed to two main reasons : first, the lack of any clear formulation of what is meant by ' generality,' and secondly, the neglect to use any rigorous technique by which alone such questions can be solved.

The terms in which the problem is commonly debated are reminiscent of a similar controversy that arose years ago in the sphere of cognitive assessment, when Spearman maintained that the results of intellectual tests were due to a single general ability, while Thorndike contended that they were highly specific. Thanks to the adoption of Pearson's method of component analysis, enabling both general and bipolar (or group) factors to be verified when present, this question seems now virtually settled. It is natural to suppose that the same technique might enable us to deal with an analogous problem in the field of social studies.

History of Factorial Investigations.—A few investigators have already attempted to apply factorial methods to the study of attitudes. But the number of published researches is surprisingly small, and (so far as our present problem is concerned) decidedly inconclusive. In this country the earliest enquiries along these lines were carried out by Burt, and dealt chiefly with children's attitudes towards different subjects of the school curriculum, books, games, and future careers (cf. (3) and references) : he inferred that the factors so discovered were the effects partly of temperamental tendencies, and partly of social influences operating chiefly through parents, teachers, and school companions.[1] Later, a succession of enquiries were carried out under his direction at the National Institute of Industrial Psychology and at the Psychological Laboratory, University College, in which similar methods were used to investigate the attitudes of adults (6). Pelling and Bulley, for example, studied the attitudes of men and women of different social classes towards the æsthetic appearance and utilitarian value of various saleable commodities (3, 4). M. Johnson (14) factorized results from a questionnaire on moral, religious, political, and economic problems distributed to 234 adults, and obtained two suggestive factors : the first was identified with the tendency towards traditional authoritative or conservative views as contrasted with dissentient or radical views, and the second with a tendency to take a realistic or rational attitude as contrasted with an emotional or sentimental attitude—factors closely resembling those already noted by Burt in the fields of personal and æsthetic preferences. Flugel and Hopkins (16) circularized more than 700 adults with a questionnaire dealing with dissentient or unorthodox views over a wide range of topics—pacifism, vegetarianism, antivaccination, euthanasia, psychoanalysis, etc. Eysenck (a former research student in the same department) factorized the results, and obtained three factors—one general and two bipolar (9). In a later enquiry (10) he has obtained

[1] I should like to express my own indebtedness to Professor Sir Cyril Burt, who has given assistance and advice, and has also permitted me to refer to and quote his unpublished *Notes on Social Attitudes.*

a general factor which he terms ' radicalism-conservatism ' and a bipolar which he identifies with James' contrast between ' tough-' and ' tender-minded ' types. In America, factorial methods have been adopted by Thurstone (23), Carlson (8), and Ferguson (11). Whereas British writers have commonly found at least one general factor, American writers usually report a number of separate group factors. The techniques employed by American writers, however, differ appreciably from those developed in this country ; and in any case, owing to the differences in social background, it would manifestly be unsafe to assume that the conclusions reached in such enquiries also hold good of Londoners, much less of the English population as a whole (cf. 19).

Objections to Factorial Procedures. Several writers have objected that, even if factorial methods are applicable to the results of cognitive abilities or mental tests, they are not appropriate to the study of emotional dispositions or of opinions and behaviour resting on judgements of value. In general, the criticisms urged follow much the same lines as those advanced in other fields of enquiry. First, it is contended that an attitude is essentially a qualitative characteristic, and as such is not amenable to quantitative treatment : here as elsewhere, we are continually told, statistical psychologists are content to publish tables and numerical analyses without giving any thought to the postulates on which all quantitative treatment must depend. Secondly, it is said, even granting the correctness of the procedure, the statistical factors found to underlie attitudes could not possibly have the real or concrete existence that is claimed for them.[1]

Such objections result largely from a misunderstanding of what the factorist is proposing to do. Arguments which are much better informed have recently been put forward by Dr. Louis Guttman in a paper discussing the relative merits of analysis by factorization and analysis by scaling (12).

Factor Analysis and Scale Analysis. Dr. Guttman doubts the validity of applying to the problems of social attitudes a procedure " borrowed straight from the field of mental testing." Such a proposal, he contends, is tantamount to " trying to treat qualitative data as though they were quantitative." His own " reluctance to attempt a fully fledged conventional factor analysis," he adds, " stems from the growing evidence of both theoretical and practical inadequacies of current formulations." " Current techniques for extracting factors usually stop at three or four, or some similarly limited number ; whereas an infinite number exists whose presence can be revealed only by an entirely different approach."

In reply it may perhaps be pointed out that, in the early stages of an enquiry, an investigator may well be satisfied if he can establish the " three or four " most important factors in the field he is studying. As regards " current formulations," it would seem that Dr. Guttman has chiefly in mind the formulations favoured by American psychologists. With much of what he says British writers would probably agree ; and there can be no question that his own procedure is likely to prove at once suggestive and fruitful, especially at later stages of the work. But the problems attacked are somewhat different ; and there seems room for both methods.

The Investigation of Structure. The object of factor analysis is not so much to discover " real or concrete entities," as to study the structure of a mass of complicated data. Burt quotes Bertrand Russell's remark : " the essence of individuality baffles description ; but the consequent perplexities may be largely avoided if the importance of *structure*, and the difficulty of getting behind structure, are realized " ((5), p. 220). By structure is meant " the scheme or pattern of relations exhibited by the members of a class, regardless of their concrete nature " (3) ; and the dispute about the relative generality or specificity of attitudes is merely the first and most obvious problem that arises in the study of their ' structure ' in this sense. Those

[1] Chambers, E. G., ' Statistics in psychology,' *Brit. J. Psych.*, XXXIII, 1943, p. 189, and Penrose, L. S., *ap.*, ' Factor analysis as a statistical technique,' *J. Roy. Stat. Soc.*, B, IV, 1950, p. 90.

writers who hold that the mind, unlike the body, is " comparatively structureless " [1] naturally favour the hypothesis that all attitudes are highly specific. Those, on the other hand, who adopt McDougall's hierarchical theory of mental structure [2] contend that mere consistency of behaviour is bound to impose some degree of generality ; and, since an individual's behaviour in certain limited fields is likely to be more highly consistent than the whole range of his behaviour, they will expect to find, not only one broad general factor, but also supplementary factors of diminishing generality. Thus, as Allport observes, " the issue involved in this lively controversy is of the greatest practical and theoretical significance, for upon its solution depends not only a proper choice of methods for the investigation of attitudes, but likewise the theory of mental organization and the structure of personality itself " (2).

II. METHODS

Methodological Assumptions. The applicability of factorial methods to qualitative data as such has been fully discussed in a paper to which Dr. Guttman does not refer.[3] However, in the investigation to be reported here, a more indirect approach was adopted, which, it is believed, may to some extent avoid the difficulties to which he and other critics refer.

A questionnaire was drawn up containing 30 statements or opinions dealing directly or indirectly with familiar political, social, or religious issues (see Appendix). To be comparable with other enquiries, a number of items were taken over from the lists drawn up by previous workers—Burt, Flugel, Johnson, Vernon, and others. Where necessary, the phrasing was modified to suit the particular procedure contemplated and (in certain cases) the current form which the issues touched upon seemed to be taking. The final list was submitted to 250 students of both sexes studying at the University or at L.C.C. classes or Extension Courses ; and each was asked to indicate the degree to which they agreed or disagreed with each statement.

In defence of this procedure [4] it may be argued that, to any proposition of the type included in our list, considered in reference to any person in the sample, there can be attached what Russell [5] has termed a ' degree of credence.'[6] This will vary with different persons, and might be defined as his impressionistic assessment of the probability that the statement in question is true or untrue. As thus defined it is a characteristic that can give rise to all the relations necessary for quantitative measurement ((5), pp. 129f.). As a result we have what Russell calls "a scale of doubtfulness." By means of such a scale we may measure each person's attitude to the idea or belief expressed in the statement submitted to him. If the person neither agrees nor disagrees, the amount of credence is evidently nil ; and can therefore be expressed by zero. If he is perfectly convinced that the statement is absolutely true or untrue, the amount can be indicated by +1·00 or − 1·00 respectively.

[1] Thomson, G. H., *The Factorial Analysis of Human Ability* (1946 ed.), pp. 311–12.
[2] See McDougall's discussion of ' affective-conative attitudes ' in *Psychology : The Study of Behaviour*, ch. III, ' The Structure of the Mind,' especially pp. 111f. ; and his diagram representing ' The Structure of Character ' in ' Supplementary Chapter VI ' of the 23rd edition of *Social Psychology* (1936), pp. 440f.
[3] Burt, C., ' The factorial analysis of qualitative data,' this *Journal*, III, 1950, pp. 166–85.
[4] The statement of methodological assumptions which follows is condensed from Burt's discussion in his notes on ' The Study of Attitudes ' (3). See also Burt, C., *The Contributions of Psychology to Social Problems*, 22nd Hobhouse Memorial Lecture, 1952.
[5] " Everything we feel inclined to believe has a ' degree of doubtfulness ' or inversely a ' degree of credibility.' . . . There is a cognate subjective conception, namely, the degree of conviction that a man feels about any of his beliefs ; but ' credibility,' as I mean it, is the degree of credence a *rational* man would give " (*Human Knowledge*, p. 359). Since it is impossible to separate men into two mutually exclusive classes, the rational and the non-rational, Burt suggests that credibility should be defined as a weighted average of the degrees of credence attached by an appropriately defined population.
[6] Strictly we ought to add " in reference to any person at any given time " ; and " *the* degree of credence " (with no qualification added) should be the average of a number of judgements given by the person on different occasions. This would lead to a consideration of the ' reliability ' of such judgements ; but that would take us too far afield in what is avowedly only a preliminary exploratory study.

In practice it proves to be at once convenient and feasible to recognize three degrees of strength in both the positive and the negative direction. Accordingly, when replying to the questionnaire, every person was requested to mark the extent of his agreement or disagreement on a 7-point bipolar scale, in which the degrees were carefully defined: (for details, see (19)).

But for our particular purpose a further set of assumptions seems required. We are proposing to investigate attitudes to a number of *different* topics or statements. If we assume that, for all these different statements, credibility or credence remains a homogeneous property, we can regard the various statements as indicating " the dimensions of a given class," and represent them by lines in one and the same multidimensional space, all passing through a common zero-point. Now to apply factorial methods we must be able to resolve an attitude into its components, and to combine these components into a resultant. If we are to adopt the methods originated by Pearson, a supplementary assumption seems necessary to which perhaps insufficient attention is commonly paid. We have to assume that the attitudes form tendencies which can be combined or resolved like forces in accordance with the parallelogram law [1]: in other words, that they can be represented by vectors which obey (at least to a first approximation) the law of vector addition.[2]

Statistical Procedure. To ensure that, for purposes of correlation, all responses should be measured in the same direction, namely, in that of the ' unorthodox ' view (as we may call it), the signs were changed for certain items, namely, those numbered 1, 3, 4, 5, 6, 7, 9, 10, 11, 14, 24, 26, 27, and 29. With sevenfold ratings either product-moment, tetrachoric, or point correlations may be calculated. However, in order that the coefficients obtained should be comparable with those computed by recent workers in the same field (as well as for other reasons) the tetrachoric method was chosen.[3] The large correlation table has been published in a previous article ((19), p. 250) ; it is therefore unnecessary to reprint it here.

The correlations were factorized by the Method of Simple Summation. Burt's 1917 formula was used. As is well known, this yields saturations similar to those obtained with Thurstone's later ' centroid method ' : the differences, which are seldom very large, are due chiefly to the fact that Thurstone determines the self-correlations or ' communalities ' on the basis of rank rather than significance, estimates them from the largest coefficient in the column, and introduces a special procedure of his own for reflecting the negative signs : he then rotates the factor-matrix to secure a ' simple structure ' of positive group factors only. Here it seemed essential to adhere to Burt's original procedure, and to obtain the self-correlations by successive approximation. Further, his method of rearranging the matrix of residuals, before reversing signs, appeared not only more illuminating, but quicker and more accurate.

[1] This, in Mill's phrase, amounts to assuming that the psychology of attitudes conforms to the ' law of the homogeneous mixture of effects.' Mill, it will be remembered, held that this assumption was *not* warranted in dealing with mental characteristics : unlike mechanical forces, which obey the principle of the ' *composition* of causes,' but like chemical properties, which obey the principle of the ' *combination* of causes,' mental characteristics (so he contends) produce new and unpredictable properties on combination (*System of Logic*, Bk. III, ch. vi). For a more rigorous definition of the two alternative assumptions, and the logical importance of the distinction, see Johnson, W. E., *Logic*, Pt. II, pp. 236f. The need for making some such assumption in applying factor analysis to subjective data is admirably seen in the oldest and simplest case of factor analysis, namely, the attempt to determine the ' primary colours.' It was an appreciation of the difficulties mentioned in the text that led the mathematician Grassmann to introduce the newer formulation of the ' laws of colour mixture,' in place of the more naïve statements dating back to Newton ; (Wundt, *Grundzüge*, II, pp. 160f.).

[2] It is not easy to verify the accuracy, or rather the approximate accuracy, of the vector law in the case of mental measurements. As Burt observes, probably the most satisfactory confirmation is the fact that it enables fairly correct predictions to be made (3).

[3] The obvious objection to the use of a tetrachoric procedure is that (as will be shown in a moment) the frequency distributions on the rating scale tend to be bimodal rather than unimodal. In defence it is usually contended that the *underlying* tendency would probably show a normal distribution. In point of fact, with distributions like the present the chief effect of using the product-moment method would be that the absolute size of the coefficients would be increased (some would say exaggerated) while their relative size would remain much the same.

III. RESULTS

Significance of Factors. To determine the significance of the factors as each was extracted, Burt's criterion was adopted : a factor cannot be regarded as significant unless it is based on a set of correlations or residuals among which more than one in 20 exceeds twice the standard error of the observed coefficient. Judged on this basis, the first four factors were found to be significant, and the fifth was not. The saturations were calculated to three decimal places ; for simplicity, however, only the first two figures are given in Table I. The items have been rearranged to indicate the classifica-

TABLE I. FACTOR-SATURATIONS

No.	Statements	Factors			
		I	II	III	IV
I*A*					
6	(No personal God)	·58	·34	·31	·29
14	(No loving God)	·60	·23	·28	·07
10	(Christ not divine)	·62	·28	·19	·19
4	(No survival after death)	·47	·47	·19	·06
12	(Good life without religion)	·50	·25	—·08	·03
8	(Sterilization approved)	·16	·37	—·45	·05
19	(Revolution approved)	·32	·05	—·42	·38
I*B*					
28	(Complete freedom of speech)	·42	·23	—·16	—·35
17	(Divorce by consent)	·43	·42	—·10	—·17
30	(Companionate marriage)	·63	·28	—·21	—·07
15	(Abortion approved)	·65	·30	—·18	—·26
21	(Birth control encouraged)	·54	·22	·08	—·26
3	(Extra-marital relations)	·63	·18	·26	—·10
1	(No divine creation)	·74	·36	·26	—·12
II*C*					
2	(Rich and poor not equal)	·27	—·06	—·24	·32
13	(Immorality of capitalism)	·69	—·31	—·25	·31
25	(Giving up of sovereignty)	·59	—·31	—·06	·32
20	(Class conflict)	·69	—·11	—·39	·30
23	(Complete nationalization)	·68	—·18	—·36	·15
7	(Socialism not bureaucratic)	·57	—·34	·12	·20
5	(C.O.s not traitors)	·40	—·39	·62	·19
29	(No corporal punishment)	·36	—·08	·09	·17
II*D*					
22	(Jews as good citizens as others)	·30	—·40	—·15	—·55
18	(Equality of sexual freedom)	·14	—·23	—·16	—·34
27	(Women not inferior)	·26	—·04	·19	—·38
26	(Jews not too powerful)	·48	—·27	·07	—·14
24	(Coloured people not inferior)	·24	—·38	·07	—·05
9	(Mixed marriages)	·50	—·30	·11	—·19
16	(Lenient treatment of criminals)	·51	—·36	·07	—·03
11	(No wide differences in intelligence) ..	·22	—·22	·27	—·12
	Contributions to variance (per cent.) ..	25·4	8·3	6·2	5·7

tion revealed by the analysis ; and a detailed examination of the table will suggest the general nature of the factors found.

First Factor. The first factor has positive saturations for all the 30 statements, and is thus a ' general ' factor. It contributes 25 per cent. to the total variance, that is, more than any other significant factor, or indeed than the other three put together. Over half the statements obtain saturations exceeding 0·50. On examining those showing the highest saturations, it would seem that this factor indicates a tendency for individuals to break away from traditional and accepted doctrines in favour of views arising out of new movements or new trends.

The generality of the factor implies that in any individual this non-compliant tendency (or the reverse) is, as it were, an habitual standpoint affecting his views in all the fields covered by the questionnaire—political, social, and religious. In a description of *Victorian England*, G. M. Young has drawn attention to a contrast in social tendencies which is very similar to the contrast here implied. " If," he says, " we range the forces operating on society at any moment into the Conformist and the Dissident, or the Stabilizing and the Exploratory, and apply this canon to the development of Victorian England, we shall remark how singularly detached they are from the traditional alignment of parties. There is one pattern of ideas and another of parties, and of the ideas neither party seems to be more or less receptive than the other " (*loc. cit.*, p. 175).

The names usually adopted by statistical psychologists—e.g., ' liberalism ' or ' radicalism '—seem unsuitable for two main reasons. First, they are primarily political terms with definite technical associations, whereas the factor is of a more general nature. Secondly, even in the field of politics they do not convey quite the same meaning as the saturations would suggest, and moreover have different meanings in American and British politics. Some term that is both ethically and politically neutral would be preferable. Professor Burt has suggested that in their general form the tendencies might be termed ' alterationism ' and ' preservationism,' and, in reference more particularly to attitudes and opinions, ' heterodoxy ' and ' orthodoxy.' [1]

Second Factor. The second factor is a bipolar factor ; about half the statements have positive saturations and half negative. It contributes just over 8 per cent. to the variance. A study of the items having positive and negative saturations suggests that it distinguishes statements representing unorthodox ethical or religious views from those representing unorthodox political or social views. It might therefore be termed a factor of *ethico-religious* versus *politico-social alterationism.* To avoid these rather clumsy designations, let us describe it more simply as referring to *personal* versus *political* topics.

Sections I and II of Table I show the nature of this division. In other connexions Burt has reported a factor very similar to this which he calls a factor for ' humane ' or ' sentimental ' versus ' efficient or realistic ' judgements. It represents a dichotomy that cuts across all party lines. Once again it is easy to find outstanding examples of the contrast in British history. In politics, for example, statesmen like Stafford, aiming at a policy of ' thorough,' become supporters of traditionalist or authoritarian views ; romantic writers like Lytton eventually reach much the same views because they favour all that is ancient both for its own sake and for the sake of the emotional associations that cluster round it. Among the opposite party, the radicalism of Bentham is of the efficient kind, that of Charles Dickens and William Morris of the sentimental kind.[2]

Fourth Factor. The third and fourth factors (second and third bipolar) contribute almost equally to the variance, namely, 6·2 and 5·7 per cent. respectively. A rough

[1] He writes : " The contrast is much wider than that between liberalism or the reverse. It is suggestive of the opposing tendencies that the old evolutionists regarded as essential to all life and growth—the tendency to *preserve* what has already persisted as contrasted with the tendency to *alter* or deviate—an opposition which they have described by phrases like ' permanence ' and ' variation,' ' anabolism ' and ' katabolism,' etc. Since the reference is here more particularly to opinions or *doxæ*, we might perhaps speak of ' palæodoxy ' and ' kainodoxy.' These terms would avoid the suggestion of praise or blame, or of particular spheres of doctrine. However, to avoid uncouth neologisms, I think we may provisionally be content with the natural terms ' orthodoxy ' and ' heterodoxy.' The prefix ' ortho- ' however, must be interpreted, not as meaning ' correct ' opinions, but opinions continuing in the same ' straight ' line as before, i.e., in the line of tradition or authority ; ' hetero- ' will then mean opinions deviating in a *different* direction. This is consistent with the dictionary definitions : e.g., for ' orthodox ' the O.E.D. gives (amongst other explanations) ' holding currently accepted opinions, not independent-minded or original.'

[2] Cf. Barker, E., *Political Thought in England : 1848–1914* (1947), especially pp. 190f.

attempt at weighted summation suggests that with continued approximation the fourth factor would emerge as the more important of the two. However, the low values of the factor-saturations as compared with their standard errors render it scarcely worth while at this stage to carry out a more laborious analysis. It seems clear that a larger sample of persons and a larger variety of questions would be needed to reach any conclusive decision on this point.[1]

The fourth factor can be readily interpreted. It divides the positive half of factor II into (A) items concerned mainly with religious topics or topics on which religious persons have strong views (e.g., sterilization) and (B) items concerned mainly with matters of personal morality, especially sex morality ; it divides the negative half of factor II into (C) items concerned mainly with relations within the group and (D) items concerned mainly with relations between groups.

Third Factor. Within these four sections the third factor may be considered to introduce a still finer sub-classification, by dichotomizing each section. However, with only 30 items each of these narrower classes necessarily contains very few items; and hence the grouping must largely depend on the accidental content of just two or three related items in each case. Any systematic interpretation is therefore bound to be somewhat speculative and artificial.

A. The first of the sections indicated by factor IV appears to be subdivided into (i) *theological* doctrines that depend directly on a religious standpoint, e.g., the existence of a personal God (4, 6, 10, 14), and (ii) *ethical* judgements of a type likely to be strongly influenced by religious preconceptions, e.g., sterilization (8, 12, 19).

B. The second is subdivided into (i) statements on personal conduct favouring a *moderate* degree of freedom, particularly in matrimonial affairs, e.g., divorce by consent (15, 17, 28, 30), and (ii) statements that would be regarded by most people as demanding an *unusual* amount of freedom, e.g., extra-marital sex-relations (1, 3, 21).

C. The third section is divided into (i) statements relating mainly to the question of *state* versus *private* (*capitalistic*) *enterprise* (2, 13, 20, 23, 25) and (ii) statements *deprecating severity* in the treatment of conscientious objectors and children, and in bureaucratic control (5, 7, 29).

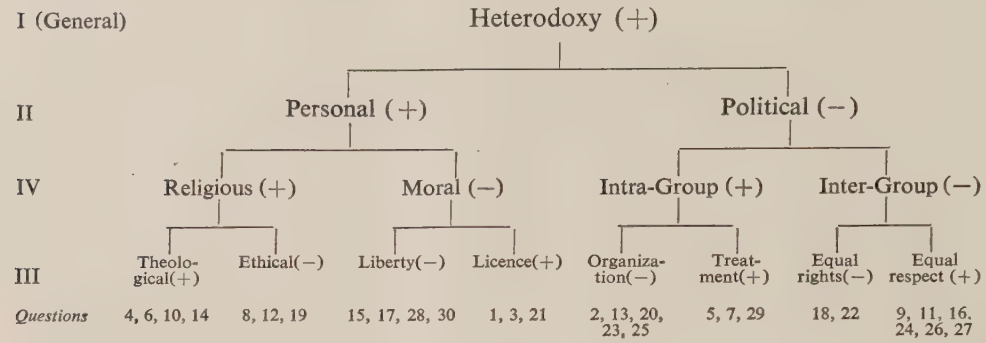
D. The last section is subdivided into (i) statements insisting on the *equality* of groups (Jews and non-Jews, women and men) (18, 22), and (ii) statements deprecating the *humiliation* of certain groups (Jews, women, coloured folk, criminals) (9, 11, 16, 24, 26, 27). The former maintain that the supposedly inferior groups have equal rights ; the latter that they deserve equal treatment. The distinction may perhaps seem a subtle one. But, as it happens, two of the statements in the questionnaire (both relating to Jews) appear among those used by Eysenck ; and Guttman draws attention to precisely the same contrast in the two sets to which, in his view, they belong. One set (which includes such statements as " Jews are as valuable as any other group "), he says, " scales a person positively if he thinks the Jews are the *same* as anybody else " ; the other (which includes such statements as " Jews have too much power and influence ") scales him positively " if he *dislikes* the Jews " ((12), p. 118). Trivial as the point may appear, it illustrates the fact that both factor analysis and scalogram analysis may lead to precisely the same distinctions.

Hierarchical Classification. It will thus be seen that, without any need for rotation, the bipolar analysis reveals a hierarchical classification and sub-classification of attitudes, dividing them into species, sub-species, and so on, the whole being covered by one generic factor. The nature of the classification will be clear from Fig. 1, where, following a device introduced by Burt, the scheme indicated by the factorial analysis is exhibited in the form of a ' genealogical tree.'

The distinctions between ' superuniverses,' ' universes,' and ' subuniverses,' to which scalogram analysis leads, are, on the surface at any rate, not unlike the distinctions here suggested where attitudes are classified by successive stages or levels according to degree of specialization. However, Guttman himself criticizes at some length any attempt to base a hierarchical classification on the results of factor analysis—an idea which he apparently attributes to Eysenck. " Even in the field of mental

[1] It is possible that part of the difficulty arises from the fact that the ordinary method of bipolar analysis assumes that the principle of subdivision introduced by any later factor is the same throughout all sections of the previous factors. Were the data based on sufficiently large samples, it might be instructive to attempt an analysis by *subdivided* bipolars, which would allow the principle of subdivision to be different for the positive and the negative sections of the previous factor (7).

FIG. 1.—*Hierarchical Scheme of Classification inferred from Factors.*



testing," he says, " there is no universal acceptance of such a hierarchical theory. . . . Thurstone has a theory of ' simple structure ' which is quite different from Eysenck's hierarchical one. . . . One can *always* fit a hierarchical theory like Eysenck's to any correlation matrix : it is a pure exercise in arithmetic " ((12), pp. 104–5). The last statement is certainly true. Take a set of correlations between 30 sets of random numbers (representing chance ratings) which have been derived from samples so small that the standard errors, and therefore the non-significant coefficients, are quite large ; then factorize them by simple summation : the result will be a succession of bipolar factors, with a sign pattern superficially resembling a hierarchical classification. But there will be two clear differences. First, the factor matrix *will not be unique :* at each stage there will, as a rule, be no one best way of reversing signs. Secondly, the factor-saturations *will not be significant.* My contention therefore is not only that we here have a hierarchical classification, but that the saturations which specify it are uniquely arranged and for the most part statistically significant.

Finally, the answer to our main problem is clear. Attitudes show both generality and varying degrees of specificity. They display a relational structure ; and that structure is hierarchical. Of the factors discovered the general factor contributes most to the variance ; but the bipolar factors (or the group factors which could be derived from them) are also indicative of genuine differences.

IV. GROUP DIFFERENCES IN THE GENERAL FACTOR

To measure the general attitude indicated by the first factor the 16 items having the highest saturations with this factor were selected. The sum of the ratings for these items, without any differential weighting, was then taken to measure each person's tendency towards an habitually heterodox or alterationist attitude.[1] The reliability of the factor-measurements thus obtained was assessed by the split-half procedure, corrected by the Spearman-Brown formula, and was found to be 0·88. This figure is high enough to indicate that these sixteen items may provisionally be used to form a reasonable scale for testing the general attitude designated by the first factor.

Table II shows the mean, standard deviation, and coefficient of variability for the two sexes and for the various political parties to which the persons questioned stated that they belonged.

Socialists have the highest factor-measurement ; Conservatives the lowest ; Liberals fall between the two, being, if anything, nearer to the Conservatives.

[1] Provided the weights are all moderately high, the use of differential weighting makes little difference to the sum obtained : cf. Burt, C., ' The influence of differential weighting,' this *Journal,* III, 1950, p. 122.

TABLE II. FACTOR-MEASUREMENTS FOR THE VARIOUS GROUPS

Group	Mean	Standard Deviation	Coefficient of Variation
Conservatives	46 ± 2·59	10·68 ± 1·83	23
Liberals	58 ± 2·87	17·00 ± 2·02	29
Socialists	76 ± 1·82	17·46 ± 1·30	22
Non-Political	57 ± 3·78	22·98 ± 2·67	40
Men	71 ± 2·78	22·38 ± 2·02	31
Women	62 ± 1·81	19·51 ± 1·28	31

The differences between the mean of the Conservative and Liberal groups and again between the mean of Socialists and those of the Conservative, Liberal, and Non-political groups were significant at the 1 per cent. level. The difference between the Conservative and Non-political groups is significant at the 5 per cent. level only. The difference between Liberal and Non-political groups is not significant.

The men are more heterodox in their attitude than the women ; and the difference is significant at he 1 per cent. level. This agrees with the results of somewhat similar enquiries carried out in America. Thus Katz and Allport (15) conclude that " the differences between the sexes increase as the degree of orthodoxy diminishes " ; Johnson (14), Stagner (21), Harris *et al.* (13), and Vetter (24) all report much the same conclusion.

Judged by the coefficient of variation, the Conservative and Socialist groups appear more homogeneous than the Liberal and Non-political group. Uniformity of response may indicate either the influence of a well-formulated and comprehensive political philosophy or else an intellectual inertia and an uncritical acceptance of well-established modes of thinking. It will be noted that a high degree of heterodoxy does not necessarily connote a high degree of original or independent thought on the part of the individual. The further relations between the general attitude thus measured and other traits of temperament and personality form a more speculative problem, and have been discussed in another paper (20).

V. DISTRIBUTION OF DEGREES OF CERTAINTY

In a suggestive study, Thouless has reported that most of his subjects, when asked to express an opinion regarding religious questions, displayed a marked bias towards certainty in either a positive or a negative direction : a suspended judgement was rare. As a result, the frequency-distribution of responses formed a U-shaped curve. On the other hand, when the same persons were asked their opinions on alleged assertions of empirical fact (e.g., " Tigers are found in China "), or on pronouncements in regard to chance occurrences, the frequency-distribution conformed to the normal curve. He argued that the causes of this difference are primarily the presence or absence of a practical implication in various statements : " in the region of action we must adopt one alternative or the other ; scepticism and doubt belong only to thought divorced from action ; and of this most people are little capable " (22).

To examine this conclusion, the responses of 183 persons (*a*) to the six statements in the questionnaire which related to religion and (*b*) to the ten statements which related to politics were selected for study, and the frequency-distributions of the averaged replies were plotted in the form of graphs.

For the statements on religious topics the distributions (with one exception) show three peaks—higher values in the middle and at either extreme. The one exception (No. 12) shows a U-shaped rather than a W-shaped curve. So far, therefore, the results do not support Thouless's finding. If we accept his general assumption that " in the region of action " most people find it difficult to adopt an attitude of neutrality or indifference, it would follow that religious doctrines are to-day no longer considered to belong to the practical field. The exception (No. 12) concerns the possibility of a good life without religion, and here the form of the statement may be regarded as bringing it within the practical sphere.

For the statements on political topics the majority of the distributions exhibit a U-shaped curve. According to Thouless's principle this result would indicate that, in the minds of those who replied, political issues *do* come within the region of conduct. However, a still more probable explanation is conceivable : in this country (except perhaps among the highly sophisticated or the highly educated) an intermediate position between the two main opposing parties has little interest or support. Burt has also suggested that, as regards political and social questions, the newspaper habitually read may further tend to " consolidate opinions into firm beliefs." Three statements (Nos. 5, 19, and 28) produced J-shaped distributions : the vast majority of those who replied were definitely of the opinion that conscientious objectors are not traitors to their country, that revolution is not necessary to progress in capitalist societies, and that unrestricted freedom of discussion is strongly desirable. F. H. Allport (1) has drawn special attention to this type of distribution, and has pointed out that " in any field of conforming behaviour, the distribution of measurable variations of that behaviour is in the shape of a steep unimodal J-shaped curve."

The occurrence of these non-normal distributions certainly raises important problems for those who desire to measure the degree of concomitant variation between attitudes in the fields here studied. In my view they do not necessarily invalidate the results achieved by the adoption of the usual correlational and factorial techniques ; but they may mean that those results can be regarded as no more than a first approximation. For methodological reasons we have provisionally assumed that the qualitative characteristics indicated by the replies are functions of some underlying quantitative variable (or variables) whose distribution is approximately normal ; but this is admittedly an assumption for which empirical evidence is desirable.[1] It is hoped in a later investigation to attempt a more intensive enquiry into the use of alternative procedures, notably that of scalogram analysis.

VI. SUMMARY

1. A questionnaire of 30 statements on social, political, and religious topics was given to 250 men and women, who expressed their agreement or disagreement with each on a 7-fold scale. The correlations were factorized by Burt's Method of Simple Summation.

2. Four significant factors were found. The first proved to be a general factor with positive saturations throughout, and was identified with a general tendency towards *heterodoxy* (' alterationism ') or the reverse. The second factor was bipolar, and appeared to distinguish *personal* from *political* problems. Of the remaining factors one appeared to divide the positive section of the first bipolar into *religious* and *moral* subsections, and the negative into *inter-group* and *intra-group* subsections ; the other appeared to introduce numerous finer subdivisions whose nature seemed to depend largely on the accidental relations in the questions selected. The whole indicates a *hierarchical classification of attitudes* by factors of increasing specialization.

3. To measure the general tendency towards heterodoxy, the sixteen items most highly saturated with the first factor were treated as forming a provisional scale. The results showed significant differences between the sexes and between the chief political groups.

4. The distributions for the replies to religious questions showed a W-shaped distribution, and for the replies to political questions a U-shaped distribution (with a few minor exceptions). It was inferred that, for those who replied, political questions have a greater practical interest than religious questions.

[1] Dr. Guttman asserts that, by ignoring the risks attending an uncritical acceptance of such assumptions, " Dr. Eysenck has arrived at nonsensical relations among his hypothetical quantities, namely, a correlation matrix which is not Gramian and partial multiple correlations greater than unity " ; as a result, he " has come a cropper in his own work " ((12), p. 108). We hope that the more conservative procedure we have adopted in calculating our correlations has escaped these graver sources of error.

REFERENCES

1. Allport, F. H. (1934). ' The J-curve hypothesis of conforming behaviour.' *J. Soc. Psychol.*, V, 141–83.
2. Allport, G. W. (1935). ' Attitudes.' *Ap. Handbook of Social Psychology* (ed. C. Murchison). London : Oxford University Press.
3. Burt, C. (1937). ' The Study of Attitudes.' (Unpublished memorandum.)
4. Burt, C. (1939). ' The factorial analysis of emotional traits.' *Char. and Pers.*, VII, 285–99.
5. Burt, C. (1940). *The Factors of the Mind.* London : University of London Press.
6. Burt, C. (1943). ' An inquiry into public opinion regarding educational reforms.' *Occup. Psychol.*, XVII, 157–67 ; XVIII, 13–23.
7. Burt, C. (1950). ' Subdivided factors.' *Brit. J. Psych., Stat. Sect.*, III, 41–63.
8. Carlson, H. B. (1934). ' Attitudes of undergraduate students.' *J. Soc. Psychol.*, V, 202–13.
9. Eysenck, H. J. (1944). ' General social attitudes.' *J. Soc. Psychol.*, XIX, 207–27.
10. Eysenck, H. J. (1947). ' Primary social attitudes.' *Int. J. Opin. Att. Res.*, I, No. 3.
11. Ferguson, L. W. (1939). ' Primary social attitudes.' *J. Psychol.*, X, 199–205.
12. Guttman, L. (1951). ' Scale analysis, factor analysis, and Dr. Eysenck.' *Int. J. Opin. Att. Res.*, V, 103–20.
13. Harris, A. J., Remmers, H. H., and Ellison, C. E. (1932). ' The relation between liberal and conservative attitude in college students and other factors.' *J. Soc. Psychol.*, III, 320–35.
14. Johnson, M. (1942). ' A preliminary study of social attitudes.' *Brit. J. Educ. Psychol.*, XII, 183.
15. Katz, D., and Allport, F. H. (1931). *Students' Attitudes.* New York : Craftsman Press.
16. Le Gulf, J., and Hopkins, P. (1944). ' A study of social and political attitudes among members of propagandist societies.' *J. Soc. Psychol.*, XX, 195–231.
17. McNemar, Q. (1946). ' Opinion attitude methodology.' *Psychol. Bull.*, XLIII, 289–374.
18. Sanai, M. (1950–51). ' An experimental study of politico-economic attitudes.' *Int. J. Opin. Att. Res.*, IV, 563–77.
19. Sanai, M. (1951). ' An experimental study of social attitudes.' *J. Soc. Psychol.*, XXXIV, 235–64.
20. Sanai, M. (1952). ' The relation between social attitudes and characteristics of personality.' *J. Soc. Psychol.* (in the press).
21. Stagner, R. (1940). ' The cross-out technique as a method in public opinion analysis.' *J. Soc. Psychol.*, II, 79–90.
22. Thouless, R. H. (1935). ' The tendency to certainty in religious belief.' *Brit. J. Psych.*, XXVI, 16–31.
23. Thurstone, L. L. (1934). ' Vectors of the mind.' *Psychol. Rev.*, XLI, 1–32.
24. Vetter, C. G. (1930). ' The measurement of social and political attitudes and related personality factors.' *J. Abn. Soc. Psychol.*, XXV, 149–89.

APPENDIX

THE QUESTIONNAIRE AS FINALLY REVISED

1.* The Universe was created by God.
2. In practice, rich and poor are not treated as equal in the eyes of the law.
3.* Except in marriage, sexual relations are morally reprehensible.
4.* Personal life continues after the death of the body.
5.* Conscientious objectors are traitors to their country and should be treated accordingly.
6.* A personal God exists.
7.* Socialism leads to too much bureaucracy.
8. Persons with grave hereditary disease should be compulsorily sterilized.
9.* Marriage between white and coloured people should be strongly discouraged.
10.* Christ was divine, wholly or partly, in a way which could not be affirmed of other men.
11.* Owing to the wide differences in innate intelligence between individuals, it is not really possible to carry out democratic principles in actual practice.
12. It is possible for anyone to live a good and virtuous life without the support of any form of religion, whether organized or unorganized.
13. Capitalism is immoral, because it exploits the worker by failing to give him the full value of his productive labour.
14.* God is a loving God.
15. Provided the life of the mother is safeguarded, abortion in the early months of pregnancy should be made legal.
16. Our methods of treating criminals are too harsh.
17. Divorce should be permitted by mutual consent, provided the needs of the children are safeguarded.
18. It is wrong that society should allow greater sexual freedom to men than to women.
19. In capitalist societies genuine social progress will be impossible without the aid of a revolution.
20. In capitalist countries the economic system forces the capitalists and workers into an irreconcilable conflict.
21. Birth control should be encouraged by putting the necessary information at the disposal of all.
22. As citizens, Jews are as worthy as any other group.
23. All large-scale means of production and distribution should be owned and operated by the State.
24.* Coloured people are innately inferior to white.
25. In the interests of peace we must give up a large part of our national sovereignty.
26.* Jews have too much power and influence in this country.
27.* Women are on the whole inferior to men in mental abilities.
28. Unrestricted freedom of discussion on every topic is desirable, in the Press, in literature, and on the stage.
29.* ' Spare the rod and spoil the child ' is a maxim that contains a good deal of truth.
30. It is right and proper that men and women should find out before marriage whether they are sexually suited (e.g., by companionate marriage).

* Marks allotted to the questions marked with an asterisk were reversed for purposes of factor analysis.

Vol. V
Part II

The British Journal of Psychology,
Statistical Section

June,
1952

THE PREDICTION OF A COMPLEX APTITUDE[1]

By CHARLES WRIGLEY
Department of Psychology, University of Illinois

I. *Introduction.* II. *The Substitution of Factor-Measurements for the Criterion-Measurements.* III. *An Arithmetical Example: the Prediction of Flying Aptitude.* IV. *A Computational Note.* V. *Summary.*

I. INTRODUCTION

The usual accounts of test validation call for a multiple correlation between the ' predictors ' (the scores for the tests) and an external ' criterion ' (cf. (11) p. 395, (12) p. 421). Thorndike and others have drawn attention to the practical difficulties which arise in personnel selection over choosing a suitable criterion (24). He urges that the criterion should be chosen on ' rational ' grounds rather than merely ' statistical ' ; it should be not only highly predictable, but also fully relevant to the occupation or work. But the tacit assumption still remains that the aptitude to be predicted is something simple or unidimensional, such as can be satisfactorily assessed by a single suitably representative measure or by an arithmetical compounding of scores into a single overall total.

In practice, however, the aptitude to be predicted is frequently complex, like flying skill or university success. And if it is multidimensional, it cannot as a rule be represented satisfactorily by any single criterion, nor can any single regression equation provide an adequate solution. In such a case we need a battery of criteria as well as a battery of predictors ; and the problem is to determine the correlational relations between the two batteries, i.e., to calculate what Pearson called a bi-multiple correlation. For doing so, probably the most familiar procedure is that employed to effect a ' canonical analysis ' (as Hotelling has called it) ((4) p. 153), or an ' external factor analysis,' as M. S. Bartlett has re-named it (cf. 20). Hotelling has discussed the procedure in some detail (18), and has developed a technique for determining the correlations of the predictors with the ' most predictable criterion ' ((17) ; (14) pp. 348–51). In that case, as he has shown, we obtain a sequence of orthogonal pairs of variates, and of correlations between them, which express invariant relations between the sets. These he calls ' canonical variates ' and ' canonical correlations.' If the number of variates is large, he suggests ((17) p. 142) a preliminary analysis of the criteria into their uncorrelated factors.

Burt has developed this line of thought further. He points out that, when we have not only several predictive tests but also several criterion measures, various methods of analysis may be adopted according to the actual nature of the psychological problem raised. He says ((6) p. 106) : " We may, for example, factorize the two sets together as though they formed a single battery ; or we may factorize each set independently, and correlate homologous factors ; or finally we may determine the maximum correlation between the two sets of variables (or their residuals), each set being combined to form a single composite variate."

Although appropriate methods have been worked out by mathematical writers, they have been rarely used by psychologists. Burt has used a form of ' canonical

[1] I should like to acknowledge the great help and stimulation I received from Professor Sir Cyril Burt during the time that I was working at University College, and again during the preparation of this paper.

analysis ' in illustrating the relations between various alternative procedures, as well as to correlate measurements for body types and temperamental types, and to compare test-data and independent criteria obtained for Army recruits. He has also given a simplified example in this *Journal* ((6) and refs.). Thomson gives an artificial example to illustrate the working procedure (23), and Vernon has also used factor analysis to obtain a more satisfactory criterion ((26) p. 111). Banks and others have used multiple criteria in analysing assessments and tests used for selecting teachers and medical students (2, 9). For the rest the vast majority of workers in the field of personnel selection have generally confined themselves to the prediction of a single criterion. Their hesitation appears to arise from two main reasons : first, the amount of arithmetical computation involved, and, secondly, the need for results that lend themselves readily to interpretation from a psychological standpoint. As Burt points out, from a psychological standpoint, the most predictable criterion is not necessarily the most relevant criterion.

The real need, therefore, is for a method of analysing the relations between the two batteries which shall be psychologically informative and at the same time practicable as regards computation. This paper describes one way of supplying this need. The suggestion may be briefly summed up as follows :

1. To reduce the battery of correlated criteria measures to a set of uncorrelated factor-measurements ;
2. To determine an orthogonal transformation, such as may yield factor-measurements having the highest multiple correlations with the battery of predictors ;
3. And finally to derive the regressional weights for the factor-measurements provided by the transformation.

The next section describes the general method. The third gives an example of the technique as applied to the prediction of success in flying training, and will argue that the results are far more informative than those furnished by a single criterion only. The fourth explains the necessary computations.

II. THE SUBSTITUTION OF FACTOR-MEASUREMENTS FOR THE CRITERION-MEASUREMENTS

From the factor analysis of the battery of criteria we may derive factor-measurements for each person in each factor (or dimension) of the complex aptitude to be predicted. The multiple correlations between the battery of predictors and each of these factor-measurement criteria may then be computed. But if we then apply an orthogonal transformation to the factor-saturations, we shall get a new set of factor-measurements and consequently of multiple correlations. Our problem is to determine what particular transformation is to be applied in order to secure the best values for the factor-measurements, keeping in mind our ultimate purpose.

We are, of course, conditioned by the procedure actually adopted for the preliminary factorization. Ideally we should choose Pearson's method of principal axes (weighted summation with self-correlations assumed to be equal to 1·00). But, in practice, when there are a large number of criterion variables (and this will usually be the case in personnel selection), it will generally be simpler to factorize the data by the method of simple summation, and to carry the analysis only so far as the first few significant factors. Here, it will be simplest to start by examining the problem with reference to the results obtained by an analysis by principal axes, and then consider the further modifications entailed by substituting approximate values reached by simple summation.[1]

Let M_t and M_c be the measurement matrices for the predictor tests and for the criteria respectively, both in unitary standard measure. Then $R_{tt} = M_t\widetilde{M_t}$, $R_{cc} = M_c\widetilde{M_c}$, and $R_{tc} = M_t\widetilde{M_c}$.

[1] The earlier part of the following argument is based on notes by Professor Burt.

Now let both correlation matrices, with units in the diagonal, be factorized by the method of principal axes. Then $R_{tt} = F_t F_t' = L_t V_t L_t'$, and $M_t = F_t P_t = L_t V_t^{\frac{1}{2}} P_t$. Similarly $R_{cc} = L_c V_c L_c'$, and $M_c = L_c V_c^{\frac{1}{2}} P_c$.

We desire to convert the principal factor-measurements P_c to canonical factor-measurements, P_x say, such that the correlation of the tests with the first factor shall be a maximum, the correlation with the second factor the maximum obtainable from the residuals, and so on. If we took the given criterion-measurements as they stand, then the familiar procedure would require us to solve the equation

$$(R_{tc} R_{cc}^{-1} R_{ct}' - \nu R_{tt})\, w = 0,$$

where w contains the required 'best weights' and ν is an undetermined multiplier. To obtain a non-trivial solution we should first solve the determinantal equation $| R_{tc} R_{cc}^{-1} R_{ct}' - \nu R_{tt} | = 0$. This determinant may be written

$$| F_t |\ | F_t^{-1} R_{tc} R_{cc}^{-1} R_{ct}' F_t'^{-1} - \nu I |\ | F_t' | = 0.$$

Since $F_t \neq 0$, we evidently have merely to find the latent roots of the symmetric matrix,

$$H \text{ (say)} = F_t^{-1} R_{tc} R_{cc}^{-1} R_{ct}' F_t'^{-1}.$$

Now instead of working with the given criterion-measurements, we propose to substitute their factor-measurements as obtained by a principal axes transformation: that is, for $R_{cc} = M_c M_c'$ we propose to substitute $\overline{R}_{cc}$ (say) $= P_c P_c'$. But the factor-measurements obtained by the transformation proposed are orthogonal: that is, $P_c P_c' = I$, and hence $\overline{R}_{cc}^{-1} = I$.

Further, $F_t^{-1} R_{tc} = F_t^{-1} M_t M_c' = F_t^{-1} . F_t P_t . M_c' = P_t M_c'$ (this last expression being simply the matrix of the correlations between the test factors and the criterion-measurements). Hence

$$H = P_t M_c' . M_c P_t' = P_t P_c' . V_c . P_c P_t', \text{ and } R_{tc} R_{cc}^{-1} R_{ct}' = F_t H F_t'.$$

The full procedure required by this formula would involve a double analysis by weighted summation and the derivation of two sets of factor-measurements. The method which I now wish to suggest is designed to lighten this labour in two ways: first, by substituting simple summation for weighted, and secondly, by restricting the factorial analysis to the criteria only.

Simple summation will account for slightly less of the variation in the criterion-measurements than weighted summation; and this slight loss in predictive power cannot be recovered. We can, however, find the most effective way of partitioning this variation, so that the greatest possible amount is included in the set of first factor-measurements, and so on. Our immediate problem therefore will be to find the orthogonal transformation of F_c which achieves this.

The procedure used in the following pages then will be to derive a set of criterion-factors by simple summation, and then, having estimated the factor-measurements ($\hat{P}_c$ say) from the battery of predictors by multiple regressional techniques, to reduce them to terms of their canonical components, so far as this is possible by the use of simple summation.

Let us write C_e for the matrix of the variances and covariances of the regressional estimates. Then $C_e = \hat{P}_c \hat{P}_c' = R_{ct}' R_{tt}^{-1} R_{tc}$. Also, if C_e is factorized, we may write $C_e = L_e V_e L_e'$, and P_e (say) $= F_e^{-1} \hat{P}_e = V_e^{-\frac{1}{2}} L_e' \hat{P}_e$. Accordingly, we may regard the values in the diagonal matrix V_e as approximations to the canonical correlations; and take P_e as the required set of criterion factor-measurements to be predicted from the test-measurement matrix M_t.

We then have $\hat{P}_c = R_{ct}' R_{tt}^{-1} M_t$ and $P_e = V_e^{-\frac{1}{2}} L_e' R_{ct}' R_{tt}^{-1} M_t$. If we do not require standardized measurements, we can take ${}_u P_e = L_e' R_{ct}' R_{tt}^{-1} M_t = L_e' \hat{P}_e = L_e' F_c' R_{cc}^{-1} M_c$.

The working method proposed, therefore, is as follows. An analysis by simple summation is first made for the criteria, and the factor-measurements computed. From the correlations between the factor-measurement criteria and the predictors, we then obtain the regression weights (B say) and the variances and covariances of the regressional estimates for the factor-measurements (the matrix C_e). Next the latent roots V and the latent vectors L of the matrix C_e are calculated. The multiplication $L'B$ gives the regression weights of the transformed factor-measurements. If the factor-measurements have been standardized, the multiple correlations are equal to the square roots of V. To interpret the nature of the criterion-factors to be predicted, we shall consequently examine, not the saturation matrix F_c, but what we may call the transformed saturation matrix $F_T = F_c L_e$.

III. AN ARITHMETICAL EXAMPLE: THE PREDICTION OF FLYING APTITUDE

Many different demands are made upon the pupil in a military flying training school. He is expected to show skill in handling an aeroplane, proficiency in theoretical and practical examinations taken on the ground, and qualities of leadership and personality which will fit him for operational duties on the completion of his training. Great difficulties have been encountered in getting a satisfactory measure of overall proficiency in training ; during the war personnel selection units were generally content to use biserial correlations with a pass-or-fail criterion.

Yet it is doubtful whether flying aptitude can be validly considered as an entity such as can be measured by a single score for each pupil. In the different Commonwealth countries, at each stage of training, ratings are made of the pupil's performance in the different branches of flying and in the various subjects of the ground examinations, and the like. Consequently there is no lack of possible criteria. The issue is to decide which to select. In these circumstances it would seem that the first step should be to make a factor analysis of the various training measurements and ratings, in order to determine how far it is possible to regard flying aptitude as a unidimensional variable or whether we are not forced to treat it as multidimensional.

In discussing problems of personnel selection Vernon has stressed ((25) p. 144) what he considers to be the limited usefulness of factor analysis from a practical standpoint ; so long as there is an external point of reference, what need is there for the selection unit to trouble about studying the internal relations of the predictive tests ? However, when we turn to the criteria, we find that, in considering aptitude for flying, we no longer have a single external point of reference. Hence it is rather in selecting suitable criterion-measurements that we find the main use for factor analysis. The problem is very like that which arises in selecting young people fresh from school in respect of their suitability for training as teachers : to make efficient schoolmasters they will require certain qualities of health, temperament, and character, as well as certain intellectual abilities. Hence, as Burt has pointed out, the appropriate criterion for such a purpose involves at least three or four uncorrelated factors, not one.

In my own investigation thirty measures of flying proficiency were available, based on performance in three different stages of flying training. These included :

(1) marks in flying tests (general flying, applied flying, instrument flying, and night flying) ;

(2) marks in the subjects of the ground examinations (navigation, airmanship, armament, meteorology, aircraft recognition, and signals) ;

(3) marks for performance in the Link trainer ;

(4) assessments of leadership ;

(5) a rating of the pupil's suitability to become an officer, made at the end of senior school training, but based on his performance throughout the course : pupils were graded in order of merit within each course, so that these ratings may be regarded as representing an overall class-mark ;

(6) the pupil's time to fly solo on each type of aircraft and a record of the number of accidents in which he had been involved.

A factor analysis by simple summation, based on a sample of 374 pupils, was accordingly carried out. This furnished the factor-saturations printed in Table I.[1] The factor-weights were next obtained ; and from them four factor-measurements were computed for each person.[2] Thus the thirty criterion measures for each individual were reduced to four factor-measurements.

[1] The table of product-moment correlations between the criteria is given in my thesis ((27) p. 156). The factor-weights, the correlations between the predictors, etc., are also to be found there, together with a fuller account of the tests and ratings in the two batteries of variables.

[2] The factor-measurements were based on the ten highest weights for each factor. The other smaller weights were regarded as zero. The computational problems will be discussed in section IV.

TABLE I. UNTRANSFORMED FACTOR-SATURATIONS: ANALYSIS OF FLYING TRAINING CRITERIA

Rating	Factor 1	2	3	4	Sum of Squares
General Flying, E.	479	432	469	−042	637
Instrument Flying, E.	365	295	384	−037	369
Night Flying, E.	287	348	357	−095	340
Leadership, E.	486	308	247	−276	468
Solo Time, E.	110	265	069	−199	127
Accidents	197	218	051	001	089
Link Trainer, E.	478	087	144	281	336
Link Trainer, S.	509	107	−058	252	337
Flying Aptitude, S.	650	418	−233	036	653
General Flying, S.	615	386	−193	053	567
Instrument Flying, S.	490	211	−146	212	351
Night Flying, S.	459	363	−171	201	412
Officer-Suitability	760	183	−317	−146	733
Leadership, S.	575	285	−420	−224	638
Solo Time, S.	171	198	−183	−016	102
Applied Flying, S.	544	−063	071	−023	305
Armament, E.	458	−369	071	−206	394
Meteorology, E.	427	−372	037	−198	361
Airmanship, E.	600	−305	−018	−063	457
Meteorology, S.	496	−293	−128	−018	348
Airmanship, S.	548	−337	−094	−282	503
Armament, S.	565	−396	−019	039	477
Signals, E.	422	−170	023	002	207
Signals, S.	369	−217	−043	024	186
Aircraft Recognition, S.	360	−219	145	068	204
Navigation, E.	528	−177	031	053	314
Navigation, S.	574	−342	024	212	493
Navigation, P.F.	279	−230	026	228	184
Mathematics, P.F.	368	−204	049	131	196
General Subjects, P.F.	565	−410	023	111	500
Sum of Squares	6·907	2·543	1·106	0·734	11·288
Contribution to Variance	23·02%	8·48%	3·69%	2·45%	37·63%

1. S. indicates senior flying school; E., elementary flying school; P.F., pre-flying training school.

2. Although the text refers to the simple summation factor-saturations, there was a rotation of $37\frac{1}{2}°$ of factors 3 and 4. The third simple summation factor was a subdivided factor (7), dichotomizing elementary-senior flying tests and theoretical-practical ground examinations. The rotation ensured that the ground examination dichotomy appeared in the fourth factor instead of the third, and was helpful in another part of my investigation. The modification makes no difference to the argument of this paper.

3. Decimal points have been omitted for the untransformed saturations.

The battery of predictors contained 12 variables: (1) an intelligence test; (2) a measure of the amount of schooling; (3) performance in the initial set of ground examinations; (4–7) four tests of spatial relations of increasing complexity; (8–10) three measures of muscular co-ordination; (11) a test of complex reaction time; (12) the age of the pupil at the time of testing.

Correlations were obtained (*a*) between predictors and each of the factor-measurements and (*b*) between the predictors themselves. From these were computed

the regressions of the factor-measurements on the predictors, and the variances and covariances of the estimated factor-measurements. The variances and covariances form the matrix C_e, and are shown in Table II. The square roots of the diagonal

TABLE II. VARIANCES AND COVARIANCES OF THE SIMPLE SUMMATION FACTOR-MEASUREMENTS, AS ESTIMATED FROM THE BATTERY OF PREDICTORS (C_e)

Factor	1	2	3	4
1	·2684	−·1472	·0785	−·0135
2	−·1472	·2029	−·0594	−·0270
3	·0785	−·0594	·0525	·0160
4	−·0135	−·0270	·0160	·1025

entries may be taken as indicating the multiple correlations with the four initial factor-measurements, namely, ·518, ·450, ·229, ·320.

The four latent roots of C_e are found to be: ·4136, ·1252, ·0653, and ·0211. The square roots—·643, ·354, ·256, ·149—represent the multiple correlations obtained with the factor-measurements when F has been transformed to give maximum-correlations with each successive factor-measurement.

The matrix L of normalized latent vectors is given in Table III. This is the orthogonal transformation matrix which is used to obtain the maximized correlations.

TABLE III. NORMALIZED LATENT VECTORS (L)

Factor	1	2	3	4
1	·7515	·4054	·4366	·2833
2	−·6037	·4228	·6745	−·0434
3	·2641	−·0855	·2299	−·9328
4	·0334	−·8059	·5491	·2187

Table IV gives F_eL, the transformed factor solution; and Table V gives the regression weights for this solution. Table V also shows the regression weights and the multiple correlation for a single criterion (such as might be used in the usual predictive study). The officer-suitability rating was chosen as the most generalized of the thirty criteria (i.e., as having the highest communality) and as intended by the Air Force authorities to give an overall verdict upon the pupil's performance.

It will be seen that the factorial method makes clear which aspects of flying aptitude are being predicted successfully, and which predictors are the effective ones, in a way that the single criterion of officer-suitability fails to do (and, indeed, in a way that a mere unweighted sum of the thirty criteria would also fail to do). Further, the existence of the regression weights enables us to identify and name the factors with a degree of confidence which we should not possess had we merely made a factor analysis of the flying training ratings independently of any reference to a predictive study.

Let us justify these assertions by considering the results in more detail. Tables IV and V indicate that there are four main aspects of aptitude for flying training:

1. *Factor I: the Intellectual Aspect* ($r = ·643$). The highest saturations are in the ground examinations, namely, General Subjects, Pre-flying, ·68; Armament, Senior, ·66; Navigation, Senior, ·65; Airmanship, Elementary, ·63; Airmanship, Senior, ·58; Armament, Elementary, ·58.

All ground examinations have saturations above ·35. The highest saturation amongst the flying tests is for Applied Flying, Senior (navigational exercises in the air), namely, ·46. No other flying test has a saturation above ·25.

TABLE IV. TRANSFORMED FACTOR-SATURATIONS: ANALYSIS OF FLYING TRAINING CRITERIA

Rating	Factor 1	2	3	4
General Flying, E.	·222	·371	·585	—·329
Instrument Flying, E.	·196	·270	·426	—·276
Night Flying, E.	·097	·309	·390	—·288
Leadership, E.	·235	·529	·325	—·166
Solo Time, E.	—·066	·312	·134	—·089
Accidents	·030	·167	·245	—·001
Link Trainer, E.	·354	—·008	·455	·059
Link Trainer, S.	·310	·054	·419	·249
Flying Aptitude, S.	·176	·431	·532	·392
General Flying, S.	·180	·386	·513	·349
Instrument Flying, S.	·209	·129	·439	·312
Night Flying, S.	·087	·192	·516	·318
Officer-Suitability	·372	·531	·302	·471
Leadership, S.	·142	·570	·224	·494
Solo Time, S.	—·040	·182	·157	·207
Applied Flying, S.	·464	·207	·199	·085
Armament, E.	·579	·190	—·146	·035
Meteorology, E.	·549	·172	—·164	·059
Airmanship, E.	·628	·167	·018	·186
Meteorology, S.	·515	·103	—·020	·269
Airmanship, S.	·581	·315	—·165	·196
Armament, S.	·659	·032	—·003	·204
Signals, E.	·426	·095	·076	·106
Signals, S.	·398	·042	·018	·160
Aircraft Recognition, S.	·444	—·014	·080	—·009
Navigation, E.	·513	·094	·148	·140
Navigation, S.	·652	—·085	·142	·202
Navigation, P.F.	·363	—·170	·098	·115
Mathematics, P.F.	·417	—·047	·106	·096
General Subjects, P.F.	·677	—·034	·031	·200
Sum of Squares	4·926	2·049	2·605	1·710
Contribution to Variance	16·42%	6·83%	8·68%	5·70%

The lowest saturations are: Time to Fly Solo, Senior, — ·04; Time to Fly Solo, Elementary, — ·07.

This factor clearly measures the intellectual aspect of the pupils' work, that is, their ability to cope with the ground examinations and the bookwork. This inference is borne out by the regression weights on the predictors. The highest are: Performance on initial examinations, ·356; Intelligence test, ·271; Amount of schooling, ·169.

It is evident that qualities of intelligence and studiousness are more important in the ground examinations and in navigational work in the air than they are in the handling of the aircraft's controls. The person who learns his bookwork well is on the average no quicker than his companions in learning to fly an aircraft solo.

From the predictive point of view, the amount of schooling appears much less important than the level of intelligence. This result confirms the wartime finding that pupils will be successful in their ground training, provided they are sufficiently intelligent and well-motivated, even though they have had only an elementary school training. It is interesting that performance in the initial examinations is a better indicator than the formal intelligence test, no doubt because different motivational tendencies, as well as different intellectual capacities, enter into such examinations.

2. *Factor II: the Leadership Aspect* ($r = ·354$). The highest saturations are: Leadership, Senior, ·57; Officer-Suitability, ·53; Leadership, Elementary, ·53; Flying Aptitude, Senior, ·43; General Flying, Senior, ·39; General Flying, Elementary, ·37. The lowest saturations are: Mathematics, Pre-flying, — ·05; Navigation, Senior, — ·08; Navigation, Pre-flying, — ·17.

TABLE V. REGRESSION WEIGHTS OF THE TRANSFORMED FACTOR-MEASUREMENTS

Predictor	Factor-Measurements 1	2	3	4	Officer-Suitability Rating
Intelligence Test	·271	—·058	—·172	—·019	·033
Amount of Schooling	·169	—·155	·141	·021	·085
Performance in Initial Examinations ..	·356	·099	—·072	—·004	·104
Spatial Relations I	—·014	·064	·042	·058	·072
II	·044	·071	·067	—·016	·102
III	·007	·051	·042	·032	·006
IV	·061	·053	·083	·013	·090
Muscular Co-ordination I	—·061	·060	—·061	—·091	·114
II	·015	·073	·077	—·064	—·032
III	—·005	·098	·037	·049	·107
Complex Reaction Time	·023	—·043	·017	—·020	·002
Age of Pupil	·083	—·137	·136	—·040	·053
Multiple Correlation	·643	·354	·256	·149	·373

The Air Force laid considerable stress on qualities of dash and vigour, especially in those intending to become fighter pilots. Skill in handling an aircraft was not enough. Youthful energy and enthusiasm were called for ; and officers in particular were required to possess these qualities (' carefree ' might be a suitable adjective to describe them). It would seem that the second factor measures those qualities of daring and enterprise which were assessed and rated highly for this purpose. The regressions are suggestive. Youthfulness appears to be an asset, since age receives a negative weight (—·137). The practical tests have positive weights : for example, the four tests depending on spatial relations all obtained positive weights, having a combined weight of ·239 ; and the measures of muscular co-ordination receive a combined weight of ·231.

Owing to the initial screening, practically all Air Force pilots were of at least average intelligence. From the figures given it will be seen that scores above average for intelligence as well as a record of longer schooling actually decrease the likelihood of securing a good rating for leadership. It would seem that, in the process of screening, superior intelligence or schooling was allowed in some measure to compensate for a lack of character qualities. Possibly for the same reason superior performance in mathematics and navigation also seems to imply a rather lower measurement for this factor, since they have negative saturations. Good work in the initial examinations, however, is favourable.

A hypothesis that might account for all these findings is the following. The qualities of youthfulness and carefreeness tend to be inimical to the steadiness and carefulness required for academic work and navigational achievement. It is the lively, practical man rather than the steady, theoretical man who is wanted as a leader. Now performance in the initial examinations is governed partly by intellectual and partly by motivational qualities. The intellectual aspects of the examinations become suppressed by the negatively weighted intelligence and schooling measurements. Hence to have done well in the initial examinations despite inferior intelligence and schooling is a sign of good motivation, so that the positive weight assigned to such examinations represents their motivational rather than their intellectual component.

Why should the spatial test scores be positively related with the ratings of leadership ? Guilford has concluded that the spatial factor is not itself an irreducible variable ((13) p. 6) : it may thus depend partly on temperament and interests. The stereotyped picture of the successful pilot is that of an extravert rather than of an introvert. And it seems reasonable to suppose that the person with good powers of visual perception may have acquired them as a result of his extraverted temperament, which would make him habitually interested in the visible world about him.

3. *Factor III : the Skill Aspect* (r = ·256). The highest saturations for this factor are : General Flying, Elementary, ·59 ; Flying Aptitude, Senior, ·53 ; Night Flying, Senior, ·52 ; General Flying, Senior, ·51 ; Link Trainer, Elementary, ·45.

This factor apparently measures skill in handling an aircraft. The Link trainer is designed to teach the skilful (i.e., manually dexterous) handling of controls in a training device on the ground. The ground examinations for the most part have zero or negative saturations : e.g., elementary meteorology, — ·15, and senior airmanship, — ·16 ; but navigation has a positive saturation : skill in handling an aircraft, of course, depends partly on navigational ability.

Positive regression weights are found, as would be expected, in the tests for spatial relations and muscular co-ordination : e.g., Spatial relations IV, ·083 ; Spatial relations II, ·067 ; Muscular co-ordination II (smoothness of arm movements), ·077.

Age here has a positive weight : it may well be that with increasing age there is an increase in steadiness of control. Although the Air Force tended to favour younger, carefree persons as fighter pilots, it preferred somewhat older, steadier men as bomber pilots. Moreover, the man who, despite lengthy schooling, is relatively less intelligent and less good in examinations is likely to be of an athletic type. This probably accounts for the positive weight for schooling and negative weights for the intelligence test and initial examinations.

4. *Factor IV : the Rate of Learning Aspect* ($r = \cdot149$). The highest positive saturations are : Leadership, Senior, ·49 ; Officer-Suitability, ·47 ; Flying Aptitude, Senior, ·39. All of the eleven highest positive saturations are ratings made at the senior school stage. The highest negative saturations are : General Flying, Elementary, — ·33 ; Night Flying, Elementary, — ·29 ; Instrument Flying, Elementary, — ·28.

All the elementary school flying tests have negative saturations. Clearly the fourth factor is a bipolar one, distinguishing elementary school performance from senior school performance—that is, distinguishing the rate of learning the new skills from the level of performance attained when the skills have been acquired. There may also be a halo effect, due to the general impression of the pupil's capacities shared by the instructors at any school. The factor is much more marked in the flying tests than in the ground examinations.

The distinction between elementary and senior variables is so definite and consistent that there is no doubt that the factor is a meaningful one. It is doubtful, however, whether it can be successfully predicted, since the multiple correlation would have to be ·20 for full significance ((11) p. 428).

Rating for Officer-Suitability ($r = \cdot373$). Table V gives the regression weights, not only for the factor-measurements, but also for the officer-suitability criterion. We are thus enabled to compare the results obtained from the use of the method of analysis advocated in this paper with those obtained by the conventional technique of correlating with a single criterion.

All but one of the regression weights for the officer-suitability criterion are positive. This is to be expected, since the rating was based on overall performance throughout training ; and in the complete picture, good spatial ability, superior muscular co-ordination, and high intelligence are all advantageous. This criterion does not make clear, however, that intelligence scores are useful mainly to predict success in ground work, and have negative weights for what we have called the ' skill ' and the ' leadership ' components of flying.

The factorial approach thus yields a much more detailed explanation why certain predictors are effective. Further, the factor analysis of the criteria makes evident the relative importance of the different aspects of flying in determining a pupil's suitability for commissioning. The saturations of this variable for the first three factors assign greatest importance to qualities of leadership (for these it has a saturation of ·53), whereas much less importance is attributed to intellectual qualities (·37) and to skill (·30).

This example also illustrates a point already stressed by Peel (21) : namely, the weakness, from the psychological point of view, of calculating just the one multiple correlation coefficient, which gives the correlation of the predictors with the best possible criterion only. At first sight the multiple correlation of ·64 with the first factor-measurements suggests that our prediction of success in flying is reasonably good. But the position appears much less attractive when it is observed that only the intellectual aspects of training are predicted to that degree, whereas the prediction of the skill and leadership components of flying is much inferior.[1]

[1] There are, of course, other reasons for the lowness of the correlations besides the inadequacies of the predictive tests : viz., the selectivity of the sample, the low reliability of most of the criteria, varying standards from one training school to another, etc. I have dealt elsewhere (27) with the effects of these difficulties upon validity coefficients, and with possible ways of surmounting them.

IV. A COMPUTATIONAL NOTE

In analysing the relations between a battery of predictors and a battery of criteria, the procedure should not only be psychologically informative ; it must also avoid any excessively laborious computation. When the number of variables is large, sums of squares and products of criteria may be obtained fairly readily by using Hollerith equipment ((15) p. 163) ; and the remaining steps required for calculating the correlation matrix are then comparatively easy. Fortunately, the larger the battery of criteria, the fewer will be the iterations required in the simple summation factor analysis, since communalities contribute proportionately less to column-sums. Once the sign pattern of the variables has been determined from the first approximation, Burt's method of 'Sum Factoring' may be used to advantage ((7) pp. 61–3).

Factor-weights may easily be obtained by the well-known formula ((22) pp. 110–12, (16) pp. 278–83), $FR^{-1} = (F'U^{-2}F + I)^{-1}F'U^{-2}$.

The major difficulty will therefore lie in obtaining the factor-measurements. The criterion measures for each person have to be condensed into factor-measurements by means of the matrix multiplication ${}_nM_c W_q = {}_nP_q$. Thus in all, $n \times c \times q$ multiplications are required ; and this may be a somewhat lengthy proceeding.

Can the Hollerith be used effectively at this stage? In general, punched-card techniques are not very suitable for matrix multiplication. The labour of forming sums of products, on the other hand, is greatly reduced by this device. This suggests an alternative procedure for calculating the correlations of predictors with factor-measurements, viz., to get the product-sum of each predictor with each criterion by using the Hollerith, and then to obtain the products of the predictors with the factor-measurements by weighting these already summed products. Thus if $p_{ij} = w_{1i}c_{1j} + w_{2i}c_{2j} \ldots + w_{ci}c_{cj}$, where p_{ij} = the factor-measurement for the ith factor for the jth person, $c_{1j} \ldots c_{cj}$ represent the scores for the c criteria for the jth person, and $w_{1i} \ldots w_{ci}$ the factor-weights for the ith factor (the β-coefficients have to be converted to b-coefficients), then the product-sum of the ith factor-measurement and the kth predictor ($k = 1 \ldots r$) is $\sum_{j=1}^{n} r_{kj}p_{ij} = w_{1i}\sum_{j=1}^{n} r_{kj}c_{1j} + w_{2i}\sum_{j=1}^{n} r_{kj}c_{2j} \ldots + w_{ci}\sum_{j=1}^{n} r_{kj}c_{cj}$. $\sum_{j=1}^{n} p_{ij}$ and $\sum_{j=1}^{n} p_{ij}^2$ may be obtained in the same way. This procedure involves more operations (namely, $n \times c \times r$ multiplications—that is, r/q times the multiplications required by the direct method, where r is generally considerably greater than q). Nevertheless it is probably more economical, because it utilizes the Hollerith.

The latent roots and latent vectors of the matrix C (which is, of course, only of order q) can readily be obtained by weighted summation ((5) pp. 320–32). In that case the columns of factor-saturations will be proportional to the latent vectors (which have to be normalized), and the sum of squares of saturations for each factor gives the latent root. Aitken (1) and Burt ((3) pp. 185–91) describe various ways of accelerating rate of convergence.

The second major difficulty in the procedure is the labour of getting the inverse of the correlation matrix of predictors (8, 10) which is required for finding the regression weights : for methods see Burt (8) and Fox (10). The difficulty is common to all multiple correlational studies involving a large number of predictors. For this reason, in the numerical example of section III the battery of predictors was restricted to twelve variables. Yet in the prediction of a complex aptitude quite a large battery of predictors may be needed, especially in view of the rather low correlations with the criteria that appear to be characteristic of even the best predictive measures in personnel selection.

Given a larger number of predictors, there appear to be two possibilities :

1. By partitioning the table into submatrices ((10) p. 112), and then using iterative methods to give greater accuracy, it may become practicable to get the inverse of a large matrix (19). Here the development of better computational aids may render help. Already punched-card equipment has been used successfully when the number of variables is greater than 20 ((15) p. 170), and in the near future high-speed electronic computers will doubtless become more generally available.

2. Alternatively, a factor analysis might be made of the battery of predictors, and factor-measurements calculated for the predictors as well as for the criteria.

V. SUMMARY

1. When an aptitude, such as flying, proves to be highly complex (i.e., multi-dimensional), it cannot be adequately represented by any single criterion. To make satisfactory predictions of such an aptitude we therefore need, not only a battery of predictors, but also a battery of criteria. A method for analysing the relations of the two batteries is consequently required which shall be at once practicable from the

computational standpoint and psychologically informative. The procedure here suggested is to reduce the battery of correlated criteria to a set of uncorrelated factor-measurements, and then predict each factor-measurement separately.

2. It is shown that the total variation in the factor-measurements as estimated from the predictors does not differ when the factors are rotated. The only effect of such a rotation is to redistribute the variation. There is, however, one particular orthogonal transformation which will yield satisfactory estimates of the uncorrelated factor-measurements and at the same time provide the highest possible multiple correlation with each successive factor. It is suggested that this should be adopted for purposes of predicting complex aptitudes, and that, for simplicity of calculation, simple rather than weighted summation should be used.

3. An example is given from the prediction of success in flying. It is shown that the regressions for predicting the transformed factor-measurements enable us to identify the factors indicated by the matrix of transformed saturations. In the present enquiry the factors were identified as : (*a*) Intellectual aspect of flying ($r = \cdot 64$) ; (*b*) Leadership aspect ($r = \cdot 35$) ; (*c*) Skill aspect ($r = \cdot 26$) ; and (*d*) Rate of learning ($r = \cdot 15$). On the other hand, the multiple correlation with a single criterion (the rating of overall performance at the time of commissioning) was only ·37 ; nor was it easy with that criterion to infer from the regressions which aspects of the pupils' performances are being estimated.

4. Computational techniques are described which will simplify the use of the procedure in studies on personnel selection.

REFERENCES

1. Aitken, A. C. (1937). ' Studies in practical mathematics. II. The evaluation of the latent roots and latent vectors of a matrix.' *Proc. Roy. Soc. Edin.*, LVII, 269–304.
2. Banks, C. (1945). *Factorial Studies : 3. Follow-up Results for Training College Students* (Ph.D. Thesis, University of London Library).
3. Burt, C. (1937). ' Methods of factor analysis with and without successive approximation.' *Brit. J. Educ. Psychol.*, VII, 172–95.
4. Burt, C. (1938). ' The unit hierarchy and its properties.' *Psychometrika*, III, 151–68.
5. Burt, C. (1940). *The Factors of the Mind.* London : University of London Press.
6. Burt, C. (1948). ' Factor analysis and canonical correlation.' *Brit. J. Psych., Stat. Sect.*, I, 95–106.
7. Burt, C. (1949). ' Subdivided factors.' *Brit. J. Psych., Stat. Sect.*, II, 41–63.
8. Burt, C. (1951). ' The numerical solution of linear equations.' *Brit. J. Psych., Stat. Sect.*, IV, 31–54.
9. Burt, C. (1948). *Memorandum on Tests and Assessments as a Means of Selecting University Entrants* (unpublished).
10. Fox, L. (1950). ' Practical methods for the solution of linear equations and the inversion of matrices.' *J. Roy. Stat. Soc.*, Ser. B, XII, 120–36.
11. Garrett, H. E. (1947). *Statistics in Psychology and Education* (3rd ed.). New York : Longmans, Green & Co.
12. Guilford, J. P. (1936). *Psychometric Methods.* New York : McGraw-Hill.
13. Guilford, J. P. (1948). ' Some lessons from aviation psychology.' *Amer. Psychologist*, III, 3–11.
14. Gulliksen, H. (1950). *Theory of Mental Tests.* New York : Wiley.
15. Hartley, H. O. (1946). ' The application of some commercial calculating machines to certain statistical calculations.' *J. Roy. Stat. Soc. Suppl.*, VIII, 154–83.
16. Holzinger, K. J., and Harman, H. H. (1941). *Factor Analysis : A Synthesis of Factorial Methods.* Chicago : University of Chicago Press.
17. Hotelling, H. (1935). ' The most predictable criterion.' *J. Educ. Psychol.*, XXVI, 139–42.
18. Hotelling, H. (1936). ' Relations between two sets of variates.' *Biometrika*, XXVIII, 321–77.
19. Hotelling, H. (1943). ' Some new methods in matrix calculation.' *Ann. Math. Statist.*, XIV, 1–34.
20. Kendall, M. G., and Babington Smith, B. (1950). ' Factor analysis.' *J. Roy. Stat. Soc.*, Ser. B, XII, 60–94.
21. Peel, E. A. (1948). ' Prediction of a complex criterion and battery reliability.' *Brit. J. Psych., Stat. Sect.*, I, 84–94.

22. Thomson, G. H. (1946). *The Factorial Analysis of Human Ability* (2nd ed.). London : University of London Press.
23. Thomson, G. H. (1947). ' The maximum correlation of two weighted batteries.' *Brit. J. Psych., Stat. Sect.*, I, 27–34.
24. Thorndike, R. L. (1949). *Personnel Selection : Test and Measurement Techniques.* New York : Wiley.
25. Vernon, P. E. (1946). ' Statistical methods in the selection of Navy and Army personnel.' *J. Roy. Stat. Soc. Suppl.*, VIII, 139–53.
26. Vernon, P. E., and Parry, J. B. (1949). *Personnel Selection in the British Forces.* London : University of London Press.
27. Wrigley, C. F. (1949). *A Psychological Study of Methods of Predicting Success in Flying Training* (Ph.D. Thesis, University of London Library).

Vol. V
Part II

The British Journal of Psychology,
Statistical Section

June,
1952

A RE-FACTORIZATION OF THE BURT-PEARSON MATRIX WITH THE ORDVAC ELECTRONIC COMPUTER[1]

By CHARLES WRIGLEY and JACK O. NEUHAUS

Department of Psychology, University of Illinois

An electronic digital computer known as the Ordvac, built on the campus at the University of Illinois for the American Army, became temporarily available for psychological work towards the close of last year. With this machine we find that it is possible, in a surprisingly short space of time, to complete a factor analysis by Pearson's method of principal axes for symmetric matrices up to an order of 24 × 24.

The machine came into commission early in December, 1951. Working methods were devised for a partial diagonalization before the end of that month ; and by February the code was finished for obtaining a full set of eigenvalues (latent roots) and eigenvectors (latent vectors or direction cosines).[2] In each case the first matrix on which the methods were tried out was what we have come to call ' the Burt-Pearson matrix.' This is a 7 × 7 matrix giving the correlations between a series of physical measurements for 2,000 criminals, obtained from Scotland Yard by Professor Karl Pearson's fellow-worker, Dr. W. R. Macdonell.[3]

The choice seemed appropriate on historical grounds. The matrix in question was the first empirical correlation matrix to be factorized by Pearson's method ; and the present occasion was the first on which an electronic computer was to be used for carrying out a complete factorial analysis by Pearson's method. Further (a point of even greater importance for our purpose) it was one of the very few matrices for which all latent roots and latent vectors were already known. We were thus able to observe the general accuracy of the computing methods to be tried out by comparing the figures secured with those already published by Burt ; and, having

[1] This study was supported in part by the United States Air Force under Contract AF33(038)–25726, monitored by the Commanding Officer, Human Resources Research Center, Attention : Director of Operations, Lackland Air Force Base, San Antonio, Texas. Permission is granted by and for the United States Government for reproduction, publication, use, and disposal in whole or in part.

[2] The code for the partial diagonalization was prepared by Dr. R. F. Clippinger and Dr. E. Cushen, that for the eigenvalues and eigenvectors by Mr. R. T. Gregory and Dr. D. J. Wheeler. Even for those with the specialized training, the coding is a difficult and lengthy piece of work ; and we wish to express our deep gratitude, not only to them, but also to Professor J. P. Nash, who is in charge of the mathematical group associated with the computer, and who has shown great interest in our problems, and has been most willing to make the codes and the machine available to us.

[3] ' On Criminal Anthropometry and the Identification of Criminals,' *Biometrika*, I, 1902, p. 202. The correlation matrix, given in the original to 5 decimal places, has been reproduced in various publications ; in this *Journal* it will be found (printed to 4 places) on p. 104 of Vol. II (Table I). The suggestion that such matrices might usefully be factorized by the method of principal axes is due to Karl Pearson. The derivation of the method is more fully discussed in Pearson's earlier paper ' On Lines and Planes of Closest Fit to Systems of Points in Space ' (*Phil. Mag.*, 6th Ser., II, pp. 559–72). Owing to the heavy computations involved, Pearson himself carried out calculations on these lines only for small 3 × 3 correlation tables. The actual factorization of the 7 × 7 matrix was first attempted by Burt, who contended that similar methods might be applied to correlation matrices obtained with psychological measurements and used it to show that very similar results could be reached by substituting the far less laborious procedure of ' simple summation ' (*Laboratory Notes on Factor Analysis*, 1927).

satisfied ourselves in regard to our general methods, we were then able in turn to check the detailed accuracy of the figures reached by Burt, whose results were obtained by procedures devised for the ordinary desk machine.

The electronic computer uses the sexadecimal system for the actual operations, but prints results in the decimal system to eight decimal places. Latent vectors are given in normalized form (i.e., as direction cosines in accordance with Pearson's original derivation), and had to be converted to factor-saturations for comparison with Burt's results. We find that the solution is accurate to at least 7 places : for example, the squared norm after diagonalization is only 0·00000003 different from its initial value ; and, on multiplying the factor-saturations to reconstruct the observed values, the re-calculated correlations never differ by more than 0·00000005 from those given in the original table.

The results finally obtained are shown in Table I opposite. In view of the labour required with a hand machine, Professor Nash, of the Mathematics Department, was much impressed by the accuracy of the figures published in Burt's article.[1] In the table giving the 49 factor-measurements, with one small exception, the differences between Burt's 4-figure values and ours are never more than 0·0002 : (the exception is in the entry against Height for factor III where his figure disagrees with ours by 0·0005). We understand that, even with a set of approximate figures as a start, the equivalent of well over a week's full time labour was required to obtain figures accurate to 3 or 4 places, i.e., in all about 60 hours' work. With the aid of the machine, the whole calculation requires about 10 minutes' machine time for a 7×7 matrix, and about half-an-hour for a 20×20 matrix. Most of the time is spent in feeding in the orders and numbers, and printing out the results. The actual computation (that is, the diagonalization of the matrix and the recording of the latent vectors) took only 60 seconds.

Three other correlation tables have been factorized in the same way, including Burt's 1915 matrix for correlations between emotional tendencies—a matrix which has recently been re-factorized by Holzinger and Harman.[2] We hope to discuss the results for the various tables more fully in a later note. Judging by these larger tables, and taking into account the time spent in scaling the matrix, punching the tape, converting the normalized latent vectors to factor-saturations, and checking the results, we estimate that from four to five hours' work altogether is required to make a factor analysis of a 20×20 matrix. In handling so large a matrix it would be laborious to multiply the factor matrix by its transpose in order to check results. We have instead sufficiently powerful checks if we find, first, that the norm and trace have remained invariant in the diagonalization of the matrix, and secondly, that the sum of the squares for each column of factor-saturations is equal to the latent root. The final results give the complete set of 20 factors ; and consequently Bartlett's test of significance of the successive latent roots (this *Journal*, III, p. 77, and IV, p. 1) may at once be applied to determine how many factors can be regarded as statistically significant.

[1] (Note by Burt.) The credit for the accuracy of the final figures should go to Dr. Banks. The early calculations were accurate only to two decimal places. A completer factorization was progressively attempted by successive research students at London Day Training College and University College who wished to practise such methods and used this matrix as providing a convenient exercise. Dr. Banks revised the figures and carried out further iterations, taking the final figures to 5 decimal places. As Dr. Wrigley points out, in our table of factor-saturations (*loc. cit.*, p. 109) there is an error in the entry against ' Foot length ' for factor VII : owing to a slip in copying, the figure printed should have been 0·0048, not 0·0489.

[2] Burt, C., ' General and Specific Factors Underlying the Primary Emotions,' *Brit. Ass. Adv. Science : Annual Report*, LXXXIV, 1916, p. 695. Holzinger, K. J., and Harman, H. H., *Factor Analysis*, 1941, p. 176.

TABLE I. FACTOR-SATURATIONS FOR PHYSICAL MEASUREMENTS

7 × 7 Correlation Matrix Factorized by Pearson's Method of Principal Axes

Measurement	I	II	III	IV	V	VI	VII
1. Head Length	·53821852	·44627677	·71189547	−·05193564	−·04039334	−·00245476	·00526434
2. Head Breadth	·41301783	·78363987	−·20623308	·41175430	·05426934	−·01645095	−·00594180
3. Face Breadth	·57505029	·62793170	−·30864731	−·41771201	−·06588083	−·01697031	·02538004
4. Foot	·87768464	−·21840873	−·04786264	·03101751	−·01710574	·42237243	−·00482586
5. Forearm	·88839436	−·33899302	−·02886713	·06901654	−·02281881	−·14067123	·26444938
6. Height	·84934660	−·21976162	−·00517781	−·05688007	·44697984	−·11484282	−·11863038
7. Finger	·85291985	−·28776059	−·05679103	·06784889	−·36010912	−·15279557	−·16990522
Sum of Squares [1]	3·79938252	1·50127259	·65096658	·36028642	·33919993	·23528687	·11360488
Latent Root [2]	3·79938260	1·50127264	·65096659	·36028643	·33919994	·23528688	·11360488

[1] Calculated from the figures in the column above.

[2] Given direct by the machine.

It may be of interest to note that the machine uses an iterative method for diagonalizing the matrix, by making successive sine-cosine transformations to reduce each off-diagonal element to zero. This is the same method as that proposed by Kelley in 1935 in *The Essential Traits of Mental Life*. It was, however, independently worked out by Goldstine about ten years later. Here as elsewhere (e.g., in Probit Analysis) we have a striking instance of the independent development or rediscovery by mathematicians of a procedure already worked out by psychologists. It might perhaps be claimed that the recent use of what are essentially factorial methods by mathematical statisticians and economists is itself a noteworthy instance in which the early statistical procedures of psychologists, after running the gauntlet of much criticism from psychologists as well as statisticians, have been adopted by later statistical writers.

We may add that the University of Illinois is building its own computer to a similar plan, and it is expected that this will be ready later this year. We then hope to develop a code for computing inverses and for making canonical analyses by the aid of the same machine.

Vol. V
Part II

The British Journal of Psychology,
Statistical Section

June,
1952

TESTS OF SIGNIFICANCE IN FACTOR ANALYSIS

By CYRIL BURT
Department of Psychology, University College, London

I. *Problem.* II. *Tests based on Analysis of Variance.* III. *Tests including Correction for Specific Factors.* IV. *Tests for Current Factorial Procedures.* V. *Direct Tests of Factor-Saturations.* VI. *Miscellaneous Methods.* VII. *Empirical Studies of Significance Tests.* VIII. *Summary and Conclusions.*

I. PROBLEM

Need for Significance Tests. The weakest feature of current factorial methods is the lack of any agreed procedure for testing the significance of the factors discovered. The standard text-books either ignore the problem or dismiss it in a few short paragraphs. Nevertheless, a closer study of articles, theses, and published researches will show that from time to time numerous alternative procedures have been proposed and tried ; and the object of this paper is to summarize the more plausible suggestions, and to indicate their advantages and limitations from both a theoretical and a practical point of view.[1]

A test of significance may be defined as " a statistical test designed to distinguish deviations due to sampling error from deviations indicating a real discrepancy between hypothesis and observation." [2] In factorial researches the processes of sampling are somewhat complex. Indeed, in any complete correlational study the psychologist may be confronted with variations in three dimensions, namely, (*A*) in the individuals measured, (*B*) in the qualities tested or assessed, and (*C*) in the times at which the measurements have been made. For brevity we can refer to them as differences between ' persons,' ' tests,' and ' occasions,' and distinguish them by the subscripts *p*, *t*, and *o*. With rare exceptions our observations must inevitably be confined to limited samples of each. Let us use *N* to denote the number in the sample of persons, *n* the number of tests, and *m* the number of occasions. Thus, the main sources of variation with the degrees of freedom to be allocated to each may be classified as in Table I.

[1] Readers will remember that, in establishing a Statistical Section of their *Journal*, the British Psychological Society desired that a certain proportion of the articles published should be " of an elementary and expository nature, suited for psychological research-workers and students possessing no advanced knowledge of mathematical statistics." The tolerance of the statistical reader is therefore asked for the frequent substitution of verbal for algebraic arguments ; a more rigorous treatment will be found in the publications cited in the footnotes or at the end of this paper. The reader, however, should be warned at the outset that, in certain factorial investigations, the testing of statistical significance may not be the chief or most important problem. The ' information ' furnished by a factor may be more instructive than the fact that is ' significant ' in the statistical sense.

[2] Mather, K., *Statistical Analysis in Biology*, 1943, glossary, p. 238 ; cf. (4), 1937 ed., pp. 335f. and 357f., (6) pp. 41f., and (20) pp. 336f.

TABLE I. CLASSIFICATION OF GENERAL SOURCES OF VARIATION WITH DEGREES OF FREEDOM AVAILABLE

Source of Variation		Degrees of Freedom
	A. Between Persons	$N - 1$
	B. Between Tests	$n - 1$
	C. Between Occasions	$m - 1$
1st Order Interactions	*AB.* Persons and Tests	$(N - 1)(n - 1)$
	AC. Persons and Occasions	$(N - 1)(m - 1)$
	BC. Tests and Occasions	$(n - 1)(m - 1)$
2nd Order Interactions	*ABC.* Residuals	$(N - 1)(n - 1)(m - 1)$
	Total	$Nnm - 1$

The complex problems that arise as a result will be clearer if we consider a concrete instance. In the earliest investigation in which the presence or absence of factors was subjected to any test of significance (3), Dr. Flugel and I collected assessments (mainly by means of tests) for $N = 43$ boys in regard to $n = 13$ cognitive processes on $m = 2$ separate occasions. In such a case we are confronted with a threefold sampling problem. Ideally, therefore, any formula for the sampling error of a factor-measurement or factor-saturation should be a function of the numbers of persons, tests, and occasions, together with their variances and covariances. It is evident, however, that, even if it could be satisfactorily deduced, the complete formula would be far too complicated for practical calculation. We shall consequently have to separate the various issues involved, and introduce certain simplified or approximate procedures. To the working psychologist this restriction has not proved so serious as it may sound. In most investigations his immediate need is simply to decide whether or not (*a*) the probability that some factor may be present is greater than (*b*) the probability that the variations on which it is based may be merely the accidental results of random fluctuations. The final acceptance of any doubtful factor will usually depend on the cumulative outcome of a succession of independent researches rather than on a single test of significance applied once for all.

The need for such a test of significance became an urgent necessity at the very outset of factorial work. Galton, to whom all correlational enquiries are ultimately due, had assumed the existence both of 'general ability' and of 'special aptitudes.' Wissler, in an early endeavour to "isolate and value general and specific ability" by studying the inter-correlations between tests, was led to suggest that "every act measured by tests is special, and their summation merely a combination conforming to the laws of chance."[1] However, no procedure was adopted for isolating 'factors' as such.

In the investigation just cited, an attempt to assess and allow for the influence of 'chance' was made in two stages. First, every test was applied twice to the same group in order to determine the reliability of each; those that proved least reliable were rejected forthwith, and, for the remainder, the measurements obtained on the two occasions were simply averaged. Then, working with these averaged measurements, and employing a procedure adapted from that put forward by Pearson for factorizing bodily measurements, the theoretical values for the inter-correlations to be expected on the hypothesis of a single common factor only were computed; and the discrepancies between the observed and the expected values were systematically compared with the relevant probable errors.

Taking the table as a whole, the majority of discrepancies proved to be no bigger than we might anticipate from the errors of sampling. But in a few outstanding cases they were disproportionately large. Accordingly, it was inferred that the presence of certain 'group' factors—for memory, motor capacity, and sensory discrimination—was more probable than not; and this inference was further confirmed by more intensive studies on larger samples.[2]

[1] 'The Correlation of Mental and Physical Tests,' *Psychol. Mon. Sup.*, III, 1901, pp. 1, 55.

[2] For example, between the teachers' assessments of intelligence and the composite test of memory (reproduction of words) the observed correlation was 0·57; the expected value only 0·23; and the probable error ± 0·08. The discrepancy was thus $4\frac{1}{2}$ times the probable error; and it was argued that by mere random sampling one would obtain so large a discrepancy only once in about 500 cases. It was because the teachers' ratings were apparently influenced by the child's ability to learn almost as much as by sheer intelligence that it seemed necessary to seek for an *internal* criterion for intelligence (defined as the 'highest common factor') instead of relying on an *external* criterion (the teachers' assessments) as previous investigators had done. Much the same evidence was obtained for a group factor underlying the tests of learning; and there were similar clusters of residuals for the motor tests and for the sensory tests ((3), Tables V and VI, pp. 161–2).

Spearman, however, in a joint paper with Hart published a year or two later, maintained that the statistical procedure thus adopted was invalid. He therefore re-examined my whole table afresh by a novel criterion—the correlation between columns. As a result he concluded that there was no evidence for what he called a ' multi-focal theory ' (i.e., the theory which postulated a number of separate ' faculties ' underlying *groups* of tests characterized by the same name—tests of discrimination, association, sensitivity, or the like) : " all the facts indicate that the correlations arise through all the performances, however different, depending on a general common factor." In a later paper he proposed a more elaborate test of significance (the ' tetrad-difference criterion ') for demonstrating the absence of any factors except the first ; this was based on the familiar determinantal criterion for a matrix of rank one, together with a formula for the probable error due to Holzinger.[1]

Thorndike, on the other hand, questioned the existence of any ' general factor.' And this has been the tendency of most American investigators. Thurstone in particular, after frankly stating that " the object of factor analysis is to discover the mental faculties," insists that all correlation tables derived from mental tests can be accounted for by a number of limited group factors without invoking any common factor running through the whole field of cognitive processes. " The presence of a general factor," he says, " *could* be indicated by that part of the communality of each test that remains unaccounted for by the common (i.e., group) factors identified in a single structure ; but, so far, we have not found any conclusive evidence for a general common factor." [2] The introduction of a general factor can, of course, nearly always be circumvented, particularly if a large number of overlapping group factors are introduced ; and the number of factors that Thurstone and his followers are usually prepared to accept is so high that several writers have started questioning their statistical significance.[3]

The need, therefore, for a rigorous method of testing significance, or at least some method whose validity will be generally acknowledged, seems undeniable.

II. TESTS BASED ON ANALYSIS OF VARIANCE

The Derivation of Factors. Directly or indirectly, all current methods of factor analysis derive from the early work on the normal multivariate distribution by Pearson and his co-workers, notably Edgeworth, Sheppard, and Yule.[4] Hence factor analysis, as a statistical procedure, may be regarded as a particular form of the ' multivariate analysis of variance.' Fisher, it will be remembered, in discussing analysis of variance, describes his purpose as an endeavour, " by a process of abstraction, to isolate causes into a number of elementary ingredients or *factors* " ; [5] and his use of the term ' factor ' suggests something more than a merely nominal similarity between factor analysis and analysis of variance. It is therefore natural to begin by asking how far the principles underlying the analysis of variance and covariance may be employed or adapted to test the significance of the factors obtained in factor analysis.

[1] Spearman, C., and Hart, B., *Brit. J. Psych.*, V, 1913, pp. 51–84. Spearman, C., and Holzinger, K. J., ' The Sampling Error in the Theory of Two Factors,' *Brit. J. Psych.*, XV, 1924, pp. 17–19 ; cf. *ibid.*, XVI, pp. 86–8 ; cf. also Pearson and Moul (7).

[2] See (9), p. vii. As Thomson and Ledermann have shown, when the average correlation exceeds $(n-2)/(n-1)$, it is impossible to avoid a general factor : (for a more general and precise formulation in terms of the largest latent root, see Thomson, G., *The Factorial Analysis of Human Ability*, pp. 267f.). But unfortunately the boundary conditions thus imposed are so high that the criterion can seldom be invoked in the case of any empirical matrix. Guilford, however, has used it with an artificial matrix to show that, if simple structure is not insisted on, a general factor may emerge with Thurstone's procedure (*Psychometrika*, VI, 1941, pp. 205–8).

[3] See Yule (4), 1912 ed., p. 252. Spearman and his collaborators have vigorously criticized this view ; but Spearman's comparison of his own ' fundamental equation ' with that of Pearson tacitly implies that his procedure may be regarded as a special case of Pearson's (cf. Spearman, C., ' Pearson's Contribution to the Theory of Two Factors,' *Brit. J. Psych.*, XIX, 1929, pp. 96f., XVIII, pp. 250f.).

[4] The modern theory of multivariate analysis may be said to start with Fisher's paper on ' The General Sampling Distribution of the Multiple Correlation Coefficient,' *Proc. Roy. Soc.*, A, CXXI, 1928, pp. 654f. Perhaps the most comprehensive survey showing the relations between different forms of multivariate analysis (including factor analysis by principal components) is that of Tintner, ' Some Formal Relations in Multivariate Analysis,' *J. Roy. Stat. Soc.*, B, XII, 1950, pp. 95f. More detailed discussions will be found in the papers by Bartlett and Geary there cited.

[5] *The Design of Experiments*, 1935, p. 100.

In considering the methods available for our purpose, whether precise or approximate, it will be helpful to recapitulate, at least in brief outline, the theoretical derivation of the procedures used by psychologists to determine 'factors,' and to note more particularly what general assumptions are made at the outset and what simplifying suppositions are afterwards introduced to secure workable procedures. Unless otherwise stated, we shall take it for granted that, as in the commonest type of problem, the investigator's aim is to factorize covariances or correlations between tests, not between persons. In that case N (the number of persons) will be greater than n (the number of tests); and no relevant information will be lost if all the measurements are reduced to terms of deviations about the average for each test (though not necessarily to terms of standard measure).

Our initial assumption will be that the matrix of observed measurements, M (say), may be regarded as composed of two portions, X, the portion due to one or more significant common factors, and Z, a residual portion due to 'error.' We may therefore begin by writing,

$$M = X + Z. \qquad \text{(i)}$$

Further, it would seem reasonable to assume that the variations in the tests that are attributable to 'error' are independent of the variations due to the common factors, or in other words that X and Z are uncorrelated, i.e., $XZ' = 0$. We do *not* at this stage assume that the 'error' variations in one test are independent of those in other tests.

Now let $A \equiv XX'$ be the $n \times n$ matrix of square-sums and product-sums 'between persons'; let $E \equiv ZZ'$ be the $n \times n$ matrix of square-sums and product-sums 'within persons'; and let $C \equiv MM' = A + E$ be the $n \times n$ total matrix of square-sums and product-sums. Then, since N is greater than n, C will be of rank n and consequently non-singular. Unless n is extremely small or N extremely large, A will in general be of a lower rank.

The precise formulation of the next step is by no means easy. Following Pearson the psychologist usually accepts it as a fundamental principle of methodology that no further factor is to be invoked until the explanatory possibilities of the factor or factors already established have been pushed to the uttermost. We shall therefore seek first for the 'highest common factor' (i.e., that which contributes most to the variance), then the next highest, and so on. In Pearson's terminology, our purpose will be to discover the fewest lines of closest fit.

Putting this requirement in algebraic form, we may say that we wish to find sets of weights, such as w (where w is an $n \times 1$ column-vector), which will yield regression equations for computing the factor-measurements, namely,

$$p \text{ (say)} = w'M; \qquad \text{(ii)}$$

and to choose these weights so that the contribution of Z shall be minimized.[1] This is equivalent to saying that we desire to minimize the ratio $\frac{w'ZZ'w}{w'MM'w}$ or (what is evidently the same thing) to maximize the ratio

$$\frac{w'XX'w}{w'MM'w} = \frac{w'Aw}{w'Cw} = \frac{w'Aw}{w'Aw + w'Ew}. \qquad \text{(iii)}$$

To do so, we may either take logarithmic differentials or, following Pearson, adopt the familiar Lagrange device. Choosing the latter as somewhat simpler, we may restate our aim by saying that we wish to maximize $w'Aw$, with $w'Cw = pp' = 1$ (say) held constant, while the elements in w' are allowed to vary. Accordingly, introducing an undetermined multiplier λ, we are led to form the expression $w'Aw + \lambda w'Ew$. We then differentiate it with respect to each of the elements in the column-vector w, and set each of the derivatives equal to zero. We thus obtain

$$w'(A - \lambda E) = 0 \qquad \text{(iv}a\text{)}$$

which may be written

$$w'E^{\frac{1}{2}}(E^{-\frac{1}{2}}AE^{-\frac{1}{2}} - \lambda I)E^{\frac{1}{2}} = 0. \qquad \text{(iv}b\text{)}$$

[1] In the main my argument will follow the principles adopted by Pearson in his paper 'On lines of closest fit' (2). In estimating constants, such as regression coefficients, Pearson was generally content to assume normal distributions and adopt the principle of least squares. His justification for such a procedure was based on the theory of inverse probability and Bayes' theorem.

Up to a point, as several statistical writers have indicated (see this *Journal*, II, p. 121 and refs.), Pearson's deduction can be restated and extended to fit the requirements of the principle of maximum likelihood, or, for that matter, to fit the methods introduced by J. Neyman, though in both cases practical difficulties emerge which can only be met by introducing certain auxiliary assumptions. However, without perhaps fully realizing the further complications involved, most psychologists have wanted, implicitly or explicitly, to keep in mind the *a priori* probabilities of the hypothetical factors which their experiments are set up to test.

I should add that the above was written before the publication of Professor Kendall's recent paper, which throws much light on the difficulties attending an adequate formulation of the initial principles required for a rigorous solution of the problems of regression and the like (28). It is comforting to find that, as he observes, " the statistician working in other fields " experiences similar difficulties, and so " can hardly afford to throw stones at the factor analyst."

A non-trivial solution implies a determinantal equation of the type

$$| A - \lambda E | \text{ or } | E^{-\frac{1}{2}}AE^{-\frac{1}{2}} - \lambda I | = 0 . \qquad \text{(v)}$$

The latter equation is known as the 'characteristic equation' of the symmetric matrix $E^{-\frac{1}{2}}AE^{-\frac{1}{2}}$, and the values for λ are called its 'characteristic' or 'latent' roots. Except with matrices of an exceptionally low order, an exact solution will be difficult to achieve with ordinary means of computation. Let us therefore introduce certain simplifying assumptions which may lead to a sufficiently close approximation.

(i) First, let us suppose that the errors of the various test-measurements are uncorrelated with each other. Then the matrix, E, will become a *diagonal* matrix of square-sums only. Let us call it U.

(ii) Next, let us assume that the variations produced by sampling in the diagonal elements of the product-sum matrix may be ignored, and that differences of scale have no objective relevance. We can then convert the product-sum matrix into the corresponding matrix of correlations, R, where all the diagonal elements will be equal, namely, 1·00. This will entail corresponding changes in U. Let $D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = R$, where $D^{-\frac{1}{2}}$ is the diagonal matrix of standard deviations; and let $D^{-\frac{1}{2}}UD^{-\frac{1}{2}} = R_u$. Then R_u will also be a diagonal matrix; and its elements will specify the ratio of the diagonal elements in U to the corresponding variances given by D. Eqn. (v) now becomes

$$| R - \lambda R_u | = | R_u^{-\frac{1}{2}}RR_u^{-\frac{1}{2}} - \lambda I | = 0 . \qquad \text{(via and } b\text{)}$$

(iii) Finally, to meet the case in which the tests have been applied on *one* occasion only, let us imagine that, on the basis of preliminary work, all the tests have been selected in accordance with the usual demand for high reliability (if possible, not less than 0·95). In that case there will be little difference between the reliabilities; and we may therefore suppose that, for purposes of a rough estimate of significance, the amount of error in each test can be presumed to be the same. In that case the values in the diagonal matrix of R_u will be equal; and we may write $R_u = uI$.

Substituting the last expression in eqn. (via) and putting $\lambda u = v$, we at last reach [1] Pearson's form of the characteristic equation ((2) p. 564, (20) p. 309), namely,

$$| R - vI | = 0 . \qquad \text{(vii)}$$

The solution of eqn. (vii) will give the latent roots of the correlation matrix (with unity in the leading diagonal); and, as Pearson has shown, these roots will specify the variances of the orthogonal factors corresponding to the 'principal axes of the correlation ellipsoid.' Having found the latent roots, $v_1, v_2, \ldots$, we can go on to find the corresponding latent vectors, $l_1, l_2, \ldots$, say, which will give values proportionate to the corresponding sets of test-weights. If these are duly normalized, we obtain $R = LVL'$ (where V is a diagonal matrix of factor-variances), $P_0 = L'M$ (where $P_0P_0' = V$), $P = W'M = V^{-\frac{1}{2}}L'M$ (where P is the matrix of standardized factor-measurements, and $PP' = I$), and $M = LV^{\frac{1}{2}}P = FP$ (where F is the matrix of factor-saturations). We thus obtain both test-weights and factor-weights.

Since $LL' = I$, the transformation $P_0 = L'M$ is orthogonal; and the matrix used to effect the rotation, L', is evidently analogous in form to that implicitly adopted in analysis of variance. This suggests the possibility of applying the methods of analysis of variance for testing the significance of the factors obtained in factor analysis. In so doing, however, we shall have to take due regard of the number of degrees of freedom on which the factor is based.[2]

Degrees of Freedom. In any $n \times n$ symmetric table of square-sums and product-sums or variances and covariances, all the figures except those in the diagonal are entered twice. If we may assume that each variance and each covariance represents a single independent observation, then (in Yule's phraseology) "the number of independent dimensions of variation" will evidently be $n + \frac{1}{2}n(n-1) = \frac{1}{2}n(n+1)$. When the first factor is deducted, certain linear constraints are imposed upon the residuals (e.g., with the simple summation formula cited above the residuals in each row necessarily add up to zero). And, according to the method of analysis employed, the cumulative number of constraints imposed by the extraction of each successive factor may easily be reckoned as follows.

[1] The foregoing is not the most direct way to reach this final equation. If we start by saying that our problem is to maximize in turn the contributions of the 1st, 2nd . . . $(n-1)$th component under the condition that they will reproduce the observed variances and covariances, we shall obtain eqn. (vii) almost immediately. But this gives no indication of the way in which 'errors' of various types are to be dealt with.

[2] For a suggestive discussion of the allocation of degrees of freedom in analysis of variance, see Cochran, W. G., 'The Distribution of Quadratic Forms in a Normal System,' *Proc. Camb. Phil. Soc.*, XXX, 1934, pp. 178–91.

(*a*) *With Variances in Diagonal.* Where Pearson's method of principal components is adopted, the variances and covariances obtained from n tests are accounted for in terms of n orthogonal factors ; and each of these is specified by a column of n saturations. Hence the factor-matrix, like the covariance or correlation matrix, will contain n^2 coefficients. But, owing to the relations obtaining between its columns, the number of degrees of freedom will be less than n^2, for each column of saturation coefficients will have a correlation (or rather a product-sum) of zero with each of the preceding columns,[1] and consequently each will contain one degree of freedom less than its predecessor. Hence the first column of saturations will represent n degrees of freedom, the second only $(n - 1)$, the third only $(n - 2)$; and so on. The total number of degrees of freedom therefore required for the n columns in the $n \times n$ matrix of factor-saturations is $n + (n - 1) + \ldots + 3 + 2 + 1$ (n terms in all) $= \frac{1}{2}n(n + 1)$. And this is the number of independent coefficients contained in the original matrix of variances and covariances.

On extracting the first factor, n degrees of freedom are lost, and the first residuals consequently contain only $\frac{1}{2}n(n + 1) - n$. On extracting the second factor, a further $(n-1)$ will be lost. And, in general, when k factors have been extracted, the total number lost will be $kn - \frac{1}{2}k(k-1)$; so that the number remaining in the matrix of residual correlations will be

$$\frac{1}{2}n(n + 1) - kn + \frac{1}{2}k(k - 1) = \frac{1}{2}(n - k)\,(n - k + 1). \qquad \text{(viii)}$$

The expression on the right shows that the numbers form the tail of the familiar series of so-called 'triangular numbers,' namely,

$$\frac{1}{2}n(n + 1),\ \frac{1}{2}n(n - 1),\ \ldots\ 15, 10, 6, 3, 1.$$

(*b*) *With Units in Diagonal.* So far we have assumed that each of the variates has its own objectively determined variance. If, however, they are expressed in standard measure, the n values in the diagonal of the correlation matrix will be fixed by this procedure ; and therefore the number of independent values in the matrix will be $\frac{1}{2}(n^2 - n) = \frac{1}{2}n(n - 1)$ only. Nevertheless, the calculation of the factor-saturations will still be based on $\frac{1}{2}n(n + 1)$ coefficients, because the self-correlations of 1·00 in the diagonal still enter into the computations. Each one of the n factor columns, therefore, must be regarded as absorbing one of these n diagonal values. Accordingly, when we extract the first factor we shall no longer be deducting n degrees of freedom from $\frac{1}{2}n(n + 1)$, but only $n - 1$ degrees from $\frac{1}{2}n(n - 1)$. And so, at each successive stage, the number now lost will always be one less than in the case of the unstandardized matrix. The last column of all will be automatically predetermined by the previous columns, and its degrees of freedom will therefore be nil. It follows that the number of degrees in the $n \times n$ factor-matrix will now be $(n - 1) + (n - 2) + \ldots 3 + 2 + 1 + 0$ (to n terms including zero) $= \frac{1}{2}n(n - 1)$; and this, as we have seen, is the number of the inter-correlations actually observed. Similarly, the number of degrees of freedom, which the successive sets of residuals represent and on which the successive factors will be based, will now be

$$\frac{1}{2}n(n-1) - nk + \frac{1}{2}k(k + 1) = \frac{1}{2}(n - k)\,(n - k - 1). \qquad \text{(ix)}$$

This again forms the tail of the series of triangular numbers, beginning one term later than before.

(*c*) *With Reduced Self-Correlations in Diagonal.* In adapting Pearson's procedure for psychological work, I suggested that we should (at any rate in certain exploratory researches) aim at extracting, not the maximum number of factors (n factors for n tests), but the minimum. This means that, instead of inserting in the diagonal cells the complete test-variance, we should insert only the variance due to the common factors ; [2] and this in turn implies that, if, by virtue of the preliminary

[1] A similar deduction holds good when simple summation is used instead of weighted summation. However, the method of obtaining the intercolumnar correlations (or rather product-sums) has then to be modified. It is evident that the calculation of any product-sum can be regarded as a device whereby the figures in the second of two correlated columns are weighted with the corresponding figures from the first, and the products then summed. Now, when factors have been obtained by simple (i.e., unweighted) summation, the intercolumnar product-sum has also to be calculated by unweighted summation. This means that the saturations in the first of the two columns to be correlated are changed to unity, while the plus or minus signs are unchanged. With this modification all the intercolumnar product-sums obtained from the factor-matrix will then still be zero. (Thus, since in the first column every sign is usually positive, the 'correlation' with this column merely yields the familiar result that each of the other columns, taken as it stands, adds up to zero ; the 'correlations' with any other column show that they also add up to zero after certain of the signs have been reversed. This second corollary provides a useful check on the arithmetic, but, as errors in published tables show, it is a check which seems unfamiliar to most computers.) From this it follows that the rule for calculating degrees of freedom must be the same for simple summation as for weighted summation.

[2] This proposal was strongly criticized at the time by Karl Pearson, and undoubtedly gives rise to special difficulties. Among mathematical statisticians it forms the commonest ground for rejecting factor analysis, or at least certain types of factor analysis (cf. G. A. Barnard, *ap.* (27), p. 91). Among psychologists Thomson has been the foremost critic of the change (cf. this *Journal*, III, p. 126).

standardization, the complete variance of each test is taken to be unity, we should substitute the sum of the squares of the common factors only, i.e., the ' reduced self-correlations.' [1] These reduced figures have to be determined from the saturations themselves. Hence the initial matrix of covariances or correlations will contain only $\frac{1}{2}n(n-1)$ degrees of freedom —the number of values actually observed. With the extraction of the first factor we now lose n degrees of freedom, with the extraction of the second another $(n-1)$, with the extraction of the third $(n-2)$, and so on,[2] just as in case (a). Hence after k factors have been extracted the number of degrees of freedom left is now

$$\tfrac{1}{2}n(n-1) - kn + \tfrac{1}{2}k(k-1) = \tfrac{1}{2}\{(n-k)^2 - (n+k)\} \qquad (x)$$

1. *Analysis of Factor-Variances based on Factor-Measurements obtained by Unweighted Summation.* The adaptation of the methods used in analysis of variance to problems arising in factor analysis was first discussed in detail in an early article in this *Journal* ; [3] and it will perhaps be helpful to illustrate the underlying principles in their simplest form by referring to the example there cited.

Four tests for the chief subjects of the scholarship examination were applied to a group of boys on two successive days. In such an experiment we have three primary sources of variation (persons, tests, and trials), three secondary sources of possible interaction, and a third order interaction available for testing significance—seven factors in all. Except for a few special enquiries, however, factorists have not been interested in the ' general factor-measurements ' for tests or subjects (which expresses differences in their relative difficulty), nor yet in the ' general factor-measurements ' for trials or occasions (which expresses the general trend of improvement due to familiarity or practice), but only in the ' general factor-measurements ' for persons, and in the bipolar or group factors due to ' interaction ' between persons and tests or (less frequently) between persons and trials. Accordingly, as noted above, in factorial studies the differences between the means for each test and between the means for each trial are nearly always eliminated at the outset by converting the initial measurements to deviations about the mean of each trial for each test (or of each test with the trials pooled). To obtain measurements for the various factors, the means for the several persons and the means or differences representing the interactions are then calculated. In a factorial analysis they will be weighted means. In an analysis of variance they will be unweighted means ; and the variation ' *between* persons ' indicated by these means will then be tested for significance by comparing it with the variation ' *within* the persons.'

Now, in an analysis of variance the calculation of the means and differences can be expressed algebraically by a matrix equation of the type $P_0 = L'M$; that is to say, the test-measurements (duly standardized) will be pre-multiplied by an orthogonal matrix, L', in which all the multipliers will now be equal, and those in the first row all positive. This is precisely analogous to the calculation of factor-measurements in factor analysis, except that in factor analysis the multipliers consist of differential weights which have been so computed as to yield in turn the ' highest common factors.' In both cases the sign pattern will, as a rule, have the same scheme of plus and minus signs : (compare

[1] Superficially this suggestion has some resemblance to that later introduced by Thurstone ((8), p. 66). The difference is that (as I have pointed out elsewhere) " my reduction of self-correlations is based on considerations not of rank but of significance." Thus, in determining the diagonal elements, Thurstone takes the number of common factors (r) to be equal to the *rank* of the side correlations (estimated in practice from the highest coefficient in each column), whereas I take the number to be equal to the *number of significant factors* (estimated by successive approximation), and thus reach a much lower figure. For this reason it is perhaps wiser to avoid Thurstone's term ' communalities ' in describing calculations, unless his method is being used.

[2] Unless the number of tests is one of the triangular numbers, the number of degrees of freedom required for the last factor will not be the same as the number of degrees of freedom remaining. In practice, however, the resulting indeterminacy does not seem seriously to affect the significance-test proposed, which is at best only approximate.

[3] ' Factor Analysis and the Analysis of Variance,' this *Journal*, I, 1947, pp. 3–27. The suggestion had previously been made at a symposium on ' Statistical Methods in Psychology ' (summarized in *Proc. Roy. Soc.*, B, CXXV, 1938, pp. 419f.), where the principle of maximum likelihood was likewise put forward as affording a possible approach : cf. also Appendix on ' The Applicability of Analysis of Variance to Educational Problems,' *Brit. J. Educ. Psychol.*, XVI, 1946, pp. 129f. As there pointed out, the apparent differences are three : in analysis of variance (i) the classification into factors is determined by the preliminary hypothesis ; (ii) the factor-measurements are treated as unweighted means ; (iii) the interest centres on problems of significance rather than of estimation. However, in my view even in factor analysis the experiment should be planned with a view to verifying or disproving the presence of factors specified by a hypothesis formulated in advance : as stated in what follows, the use of differential weights makes very little difference to the main factors ; and in the type of problem considered here the question of significance is paramount.

matrix of test-weights in Tables III and XIII, *loc. cit.*, pp. 7 and 23). But, as I have shown elsewhere ((26) p. 122) for the first two or three factors at any rate the use of differential weighting makes very little difference.[1] Hence the hypothetical measurements tested in an analysis of variance can be regarded as approximate assessments for the same set of factors as are obtained in an ordinary factorial analysis.

On these grounds it was suggested that a plausible procedure would be to carry out a double analysis,[2] first, a provisional analysis into factors to determine the sign-pattern of the test-weights, and secondly, a significance-test applied to the corresponding factors as calculated by an analysis of variance. (Where, as in the instance given, possible patterns are obvious in advance, it may be better to reverse the order of calculation.) Evidently, if the first or 'general' factor, obtained *without* weighting, discriminates significantly between the individuals tested, it follows *a fortiori* that the same factor, when calculated with the improved weights, will be statistically significant, since weighting must increase the variance for that factor, and so diminish the residual variance (23). On the other hand, the later factors may be diminished by the process of weighting, and show smaller resemblance with those obtained by analysis of variance. Hence with these the procedure may be less satisfactory.

TABLE II. ANALYSIS OF VARIANCE: UNWEIGHTED SUMS

Factor	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio	Probability
I	A. Persons	4	3·0810	0·7702	21·94	0·01–0·001
II	AB. Persons and Subjects	4	0·6257	0·1564	4·46	0·05–0·20
III	AC. Persons and Days ..	4	0·1520	0·0380	1·08	> 0·20
IV	ABC. Residual	4	0·1413	0·0351	—	—
	Total	16	4·0000	—	—	—

As we have seen, with the illustrative experiment just cited, there should in theory be three primary sources of variance (*A*, persons, *B*, ability tested, *C*, time of testing), three interactions of the first order (*AB*, *AC*, *BC*), and one of the second order (*ABC*). In the ordinary type of factor analysis, figures for *B*, *C*, and *BC* disappear.[3] Here therefore we are left with four factors only. Our aim will be to test the first three with the aid of the fourth. Table II gives the simplified analysis for (*A*) the general factor for persons, (*AB*) the bipolar factor classifying tests according to type of subject, and (*AC*) the bipolar factor classifying tests according to time of day: the fourth component (*ABC*), the only other component available with a 4 × 4 table, is the second order interaction, used as a 'residual.' It will be seen that the first factor (*A*) is significant at the 5 per cent. level, the second (*AB*) barely significant at that level, and the third (*AC*) definitely non-significant.[4]

2. *Analysis of Variance based on Factor-Measurements obtained by Weighted Summation.* Now, when we proceed to factorize the same set of test-measurements by Pearson's method of principal axes ((22) p. 19), the factor-saturations and factor-measurements are computed by weighted summation. As factors so obtained are entirely independent, there would seem no theoretical reason why the same procedure should not still be applied. However, since the best fitting weights are calculated from the actual data, the degrees of freedom will be changed.

[1] Professor Kendall, in using Pearson's method of principal axes to factorize agricultural data (*J. Roy. Stat. Soc.*, CII, 1939, p. 44), has made much the same observation.

[2] The principle of carrying out a separate, prior, and simpler test for significance before proceeding to make a more precise estimation was often adopted by Karl Pearson (cf., for instance, *Tables for Statisticians*, 1914, p. xxxvi).

[3] In a factorial analysis used to investigate reliability, these sources of 'error' may acquire special importance. Thus, if we take the unstandardized measurements (i.e., the crude marks), we can also test the differences between the test-means, that is, the variation of the tests in difficulty; and in the present case this proves to be unquestionably significant: cf. *loc. cit.*, I, p. 10, Table IV, and *Laboratory Notes on the Estimation of Reliability*.

[4] In the original investigation 252 boys were examined, and the examination-papers subdivided into eight tests. However, to simplify the illustration it was limited to six *persons* tested with four tests: the tests related to two different *types of subject*, 'intellectual' and 'practical,' and were applied on two different *days*. (When standard measure is used two constants for each—the mean and the standard deviation—are derived from the data; hence in each case two degrees of freedom are lost.)

TABLE III. ANALYSIS OF VARIANCE: WEIGHTED SUMS

Factor	Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio	Probability
I	A. Persons	7	3·1021	0·4432	4·44	0·01–0·05
II	AB. Persons and Subjects	5	0·7328	0·1466	(3·55)	0·05–0·20
III	AC. Persons and Days ..	3	0·1593	0·0531	(9·16)	0·05–0·20
IV	ABC. Residual	1	0·0058	0·0058	—	—
	Total	16	4·0000	—	—	—

The procedure suggested [1] is illustrated for the same set of marks in Table III. The sums of the squares are taken from the last column of the centre panel of Table IV in the article just quoted. When one of the smaller factors proves to be non-significant (factor III in the present case) the sum of squares derived from this factor can be pooled with that of factor IV; we shall then find that a further pooling is desirable for factor II. The conclusions ultimately reached are practically the same as before.

The procedure just described is primarily of theoretical interest; and, though it has occasionally been used in actual practice, it suffers from at least two drawbacks: (i) the calculations required are long and tedious; (ii) the methods adopted for allocating the degrees of freedom have not escaped criticism, and the critics themselves are not wholly in agreement.

With the methods so far discussed the test of significance turns as usual on *the ratio of two variances*. The variances that we ultimately desire to compare are (i) a variance determined by hypothesis and (ii) a variance calculated from the observed data. Now if the comparison could be so framed that the value in the denominator is a theoretical variance fixed by hypothesis appears in the denominator, then the appropriate test would be obtained by calculating chi-squared, since that is the form of the variance ratio suited to the case in which one of the variances is fixed this way. And the procedure would then be greatly simplified. Let us therefore enquire whether some practicable test, approximate if not exact, can be developed along these lines.

3. *Chi-Squared Tests.* Owing to the heavy calculations entailed by Pearson's method of principal components, a simpler mode of analysis was put forward from the very start; and (except for a small number of researches which owe their inspiration to Hotelling or Kelley) Pearson's complete method has in practice seldom been used. When first adapting his general idea for problems in psychology I ventured to suggest a shorter formula based on one which he himself had derived for fitting a manifold contingency table. If s_i denotes the sum of the coefficients in the ith row of the table, s_j the sum of those in the jth column, and S the grand total, then the expected value for the i–jth cell is given by

$$\hat{r}_{ij} = \frac{s_i s_j}{S} = \frac{\Sigma r_{ik} \Sigma r_{jk}}{\Sigma\Sigma r_{kk'}} \quad (k, k' = 1, 2, \ldots n). \qquad \text{(xi)}$$

In factorial work the problem of significance is essentially a problem of 'goodness of fit.' To obtain his formulæ for the 'lines of closest fit' Pearson adopted the principle of least squares; and, with his procedure, it is easy to show that the factors which give the closest fit to the $n \times N$ table of observed test-*measurements* (expressed in standard measure) are also the factors which give the closest fit to the $n \times n$ table of test-*correlations*, with unity in the diagonal. Hence the relative closeness of fit provided by different factors can be judged by examining the sum of the squares of the residual

[1] The degrees of freedom are assigned according to the principles explained in the article just cited. Professor Bartlett would apparently neglect the effect of fixing the standard deviation: in the above example (Table III) his allocation of the degrees of freedom would therefore be $8 + 6 + 4 + 2 = 20$ (cf. (22), p. 24, and written communication). However, since this paper was written, he has published a fuller discussion of the various problems thus raised (29).

correlations. On fitting the correlations by means of the simplified formula given in eqn. (x), it seemed natural to test both goodness of fit and statistical significance by the supplementary procedure introduced by Pearson for dealing with contingency tables, and to calculate a quantity analogous to what he would have termed the square contingency [1] ((20), ch. xxx). We might expect this quantity to have the form

$$\chi^2 = N\Sigma\Sigma d_{ij}^2, \qquad \text{(xii)}$$

where d_{ij} denotes the discrepancy between the observed and theoretical values in the ith row and jth column of the correlation matrix.[2]

In practice, when tried with artificially constructed tables, this procedure gave satisfactory results provided the observed correlations were not much greater than 0·50 and N well over 30. With larger coefficients it is less trustworthy ; and an attempt to offer a theoretical basis for such a test indicates that the assumptions required for simplifying the formula in this way are far from being adequately fulfilled. To begin with, the constants required to specify the hypothesis have to be calculated from the data : the substitution of χ^2 is consequently rather like using a table appropriate to d/σ when all we know is d/s. Secondly, the distribution of the correlation coefficient is not normal, and with large values the departure from normality is considerable even when N is large. Thirdly, in the $n \times n$ table of correlations based on n variates the errors of the coefficients are themselves correlated (1) ; and a significance test which ignores this fact can hardly be regarded as very precise. As we shall see, the introduction of minor adjustments in the equation given above may to some extent compensate for these shortcomings. But, before considering the possibility of correction along these lines, it will be helpful to consider the results of a more rigorous examination of the problem. Recent developments in the general theory of multivariate analysis (for which see (21), ch. xi, or (19), II, ch. xxviii) provide such an approach and may seem to offer a reasonable prospect of success.

4. *Testing Factor-Variances by the Methods of Multivariate Analysis.* The psychological student, whose knowledge of mathematical statistics is likely to be limited, may find it easier to appreciate the equations that arise if he begins by noting that the part played in *univariate* analysis by the *variance* or standard deviation is in *multivariate* analysis taken over by the *determinant* of the dispersion matrix, namely,

$$|R| = \begin{vmatrix} \sigma_1^2 & \sigma_1\sigma_2 r_{12} & \cdots & \sigma_1\sigma_n r_{1n} \\ \sigma_2\sigma_1 r_{21} & \sigma_2^2 & \cdots & \sigma_2\sigma_n r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_n\sigma_1 r_{n1} & \sigma_n\sigma_2 r_{n2} & \cdots & \sigma_n^2 r_{nn} \end{vmatrix}.$$

Evidently, if $n = 1$, then $|R| = \sigma_1^2$, and if $n = 2$, then $|R| = \sigma_1^2\sigma_2^2(1 - r_{12}^2)$, expressions which will be already familiar in their non-determinantal form. On this account $|R|$ has been called by Wilks [3] the ' generalized variance ' of the distribution. Geometrically the concentration of a given multidimensional distribution about its centre of gravity may be conveniently represented by what is termed the ' ellipsoid of concentration.' The ratio between the volume of such an ellipsoid and its maximum value is given by the determinant in standard measure and is called by Frisch [4] the ' coefficient of scatter ' : (cf. (20), pp. 301, 406).

Now, if we take $\sigma_i^2 = 1$, and then factorize R by the method of principal axes, we have $|R| = v_1 v_2 \ldots v_n$, where v_i denotes the ith latent root or factor-variance ; and by the ordinary formula for the volume of an ellipsoid the volume of the concentration ellipsoid will be proportional to the product of those roots. In the paper already cited, Wilks suggested a criterion for testing the

[1] Spearman and others objected to the method because it " borrowed a test for fit developed for use with *discontinuous* frequencies and used it for deviations in coefficients of correlation which are essentially *continuously* graded quantities " ; and a similar difficulty is often raised by the student of psychology. In point of fact, the criticism cuts in exactly the opposite direction. The chi-squared distribution is defined primarily as " the distribution of the sum of the squares of the deviations in a continuous quantity from a mean." Hence it is the application of this continuous distribution to discontinuous frequencies that requires defence, not vice versa (cf. (20), p. 233 ; (6), 5th ed., p. 96).

[2] If chi-squared is calculated from the square-sums and product-sums before averaging (i.e., before division by N), the multiplier n will naturally be omitted from eqn. (xi).

[3] ' Certain Generalizations in the Analysis of Variance,' *Biometrika*, XXIV, 1932, pp. 471f.

[4] *Statistical Confluence Analysis by Means of Complete Regression Systems*, 1934, pp. 14f.

significance of the correlation matrix R by a function based on the correlation determinant $|R|$ Thus, if we could obtain an expression, such as

$$\chi_n^2 = -N_n' \log_e |R| = -N' \log_e v_1 v_2 \ldots v_n,$$

for testing the significance of the entire matrix, we might hope to derive a subsidiary test for the factors remaining after k factors have been extracted either by partitioning χ_n^2 into additive components or by deducing an expression for χ_{n-k}^2 in terms of the relative values of the roots or their products.

The sampling theory of the latent roots of symmetric matrices like R has now been worked out.[1] But it leads to somewhat complicated equations which are scarcely practicable for purposes of calculation. However, by using the relation between the limiting form of a likelihood function and the χ^2 distribution, Bartlett [2] has developed a convenient approximation, namely,

$$\chi^2 = -(N - n - \tfrac{1}{2}) \log_e R_q, \qquad \text{(xiii)}$$

where $(n - k) \equiv q$ denotes the number of factors still to be extracted, and

$$R_q = q^q \frac{v_{k+1}\, v_{k+2} \ldots v_n}{(v_{k+1} + v_{k+2} + \ldots + v_n)q}$$

$$= q\text{th power of } \frac{\text{Geom. mean of last } q \text{ roots}}{\text{Arith. mean of last } q \text{ roots}}.$$

The number of degrees of freedom are given by eqn. (ix) above.[3]

Example 1. The method of calculation may be illustrated by applying it to a correlation matrix obtained with 5 variables and 80 cases. This gives $N - n - \frac{1}{2} = 80 - 5 - \frac{1}{2} = 74{\cdot}5$. Cols. (v) to (vii) of Table IV show the method of testing the significance of the total residual correlation left after k factors have been extracted ; cols. (viii) to (x) that of each factor separately after all have been extracted. The second set of values for χ^2 are derived from the first by successive subtraction, and the first set of degrees of freedom from the second by successive addition.

TABLE IV. CHI-SQUARED ANALYSIS (EXAMPLE 1)

(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
k	Latent Root (v)	R_q	log R	χ^2 (74·5 log R_q)	D.F.	Border-line	χ^2 (Diffs.)	D.F.	Border-line
0	2·443	0·287	−1·248	93·0	10	18·3	66·8	4	9·5
1	0·957	0·704	−0·351	26·2	6	12·6	10·2	3	7·8
2	0·834	0·807	−0·214	16·0	3	7·8	15·8	2	6·0
3	0·406	0·997	−0·003	0·2	1	3·8	0·2	1	3·8
4	0·360	1·000	−0·000	0·0	0	—	—	—	—
Total	5·000	—	—	—	—	—	93·0	10	—

[1] Cf. Hsu (14) and Wilks (21), sect. 11.11, 'The Sampling Theory of the Roots of Certain Determinantal Equations.'

[2] A chi-squared approximation for testing the significance of the product of *all* the latent roots of a correlation matrix was given by G. E. P. Box in a discussion of variance ratio approximations for certain likelihood criteria (*Biometrika*, XXXVI, 1949, pp. 344f). Bartlett's discussion proceeds to the further problem, which is more important in factor analysis, namely, the derivation of a practicable test for the separate and successive factors ((25), pp. 77–85). In eqn. (ix) I give a version of Bartlett's formula which is algebraically identical with his, but seems to lend itself more readily to actual computation, and is I fancy more easily understood. This form of the equation and worked example are taken from (26), p. 118. The expression in the denominator arises from the need to re-scale the values of the roots to unit mean variance as each such factor is extracted.

[3] In his more recent paper, Professor Bartlett considers it " safer to allocate the number appropriate to the unstandardized case, namely, $\frac{1}{2}(n - k + 2)(n - k - 1)$ " ((29), p. 341).

Example 2. Bartlett himself illustrates [1] his formula by applying it to the example I had used to illustrate analysis of variance (discussed above, pp. 116–17). His table and calculations are reproduced below.

TABLE V. CHI-SQUARED ANALYSIS (EXAMPLE 2)

Latent Roots	D.F.	χ^2	Value for $P = 0{\cdot}05$
$\nu_1\ \nu_2\ \nu_3$	$3 + 2 + 1$	$-\ 2{\cdot}833$ (log $0{\cdot}0021$) $= 17{\cdot}47$	12·59
$\nu_2\ \nu_3$	$2 + 1$	$-\ 2{\cdot}167$ (log $0{\cdot}025$) $=\ 7{\cdot}97$	7·82
ν_3	1	$-\ 1{\cdot}5$ (log $0{\cdot}136$) $=\ 3{\cdot}00$	3·84

$R_2 = 0{\cdot}159 \times 0{\cdot}006 \times (2/0{\cdot}165)^2 = 0{\cdot}136$;
$R_3 = (0{\cdot}0021/3{\cdot}102) \times (3/0{\cdot}898)^3 = 0{\cdot}025$;
$R_4 = |\ R\ | = 3{\cdot}102 \times 0{\cdot}733 \times 0{\cdot}159 \times 0{\cdot}006 = 0{\cdot}0021.$

This procedure seems by far the best for the cases it is designed to meet. However, with all the foregoing methods no allowance, it will be seen, is made for ' individual ' or ' specific ' factors as such. Now, in psychological work, many of the ' specific ' factors, if they could be effectively isolated, would prove to be fully significant. And with the formula given above, they are incorporated into the equations for the common factors. Hence the procedure proposed usually tends to magnify the apparent number of significant common factors. Indeed, the result will often be that the test of significance pronounces that all the principal components are significant. It is therefore plainly desirable to enquire whether the methods so far discussed cannot be modified to allow for the presence of ' specific ' factors, or, failing that, whether new methods can be devised.

III. TESTS INCLUDING CORRECTION FOR SPECIFIC FACTORS

The Four-Factor Theory. On *a priori* grounds I suggested in an earlier paper that factors may be logically divided first into (i) ' common ' (or inclusive) factors and (ii) ' non-common ' (or exclusive) factors ; we may then subdivide the former into (ia) ' general ' factors and (ib) ' supplementary ' (i.e., ' bipolar ' or ' group ') factors, and the latter into (iia) ' individual ' factors and (iib) ' error ' factors. This yields what has been termed a ' four-factor theory ' ((15) p. 109), i.e., a theory of four *kinds* of factor.[2] Now with our present problem we are not concerned with the subdivision of common factors : these we may therefore leave merged into a single group. But to allow for the presence of ' individual factors,' i.e., non-chance factors each peculiar to a single test, we must introduce another term into our initial matrix eqn. (i), and write,

$$M = X + Y + Z, \qquad \text{(xiv)}$$

where X (as before) denotes the contribution of the ' common ' factors, Y that of the ' individual ' factors, and Z the residual portion due to ' error.' Since the ' individual ' (or ' specific ') factors are uncorrelated, we shall have $YY' = B$, a diagonal matrix. For convenience in simplifying our formulæ, we may further suppose that the correlations between the errors may be disregarded. We can then assume that ZZ' will also be a diagonal matrix.

[1] *J. Roy. Stat. Soc.*, B, XII, 1950, p. 86 ; cf. (25), p. 79. Since the number of cases is here so small, chi-squared is calculated by taking, instead of $(N - n - \frac{1}{2})$, a more exact multiplier which allows for the increasing number of factors already extracted, viz., $N - 1 - 1/6(2n + 5) - 2/3k$.

[2] This view was accepted by the Brit. Ass. Committee on ' Mental and Physical Factors in Education ' : cf. also *Report on Psychological Tests of Educable Capacity*, 1924, pp. 24f., and (for a fuller account) *Marks of Examiners*, 1937, pp. 25f.

1. *The Reliability of Factors.* The only way to distinguish Z from $X + Y$ will be to use some objective procedure for assessing what is commonly called the 'reliability' of each test, e.g., (as was usually done in early investigations) to apply each test twice. In that case the familiar formula for calculating the correlation between two weighted sums will enable us to determine a reliability coefficient for each factor: we shall have in fact

$$r_{jj} = w_j' R w_j, \qquad \text{(xv)}$$

where w_j' are the test weights for computing measurements for the jth factor (in unitary standard measure) by weighting the observed test-measurements (averaged as above described), and R the correlation matrix with the observed reliability coefficients in the leading diagonal.[1]

If there is no means of assessing reliability, then it will be impossible to separate ZZ' and YY', and we shall have to work with $U = YY' + ZZ'$. When the measurements are in standard measure, U will become R_u; and we shall have $R = R_c + R_u$ as before.

2. *Correction for Specificity.* Now we are not, as a rule, interested in the 'individual factors' for their own sake; and their presence is usually regarded as diminishing the precision of the tests selected by introducing irrelevant influences.

For instance, in a battery of tests of educational ability, we might include tests of spelling and reading, to measure (i) jointly with other tests, the general factor of 'intelligence' or 'intellectual ability' and (ii) the supplementary 'verbal factor.' In that case any factor peculiar to (say) spelling as such would manifestly reduce the value of the spelling test for these special purposes. (If we desired to investigate the presence of this *third* factor, we should include at least *two* spelling tests, thus turning the 'individual' factor into a 'group factor'). The presence of 'individual' factors therefore still further increases the unreliability of our measurements.

Accordingly, with k common factors, $u_i^2 = 1 - r_{ig}^2 - r_{ih}^2 - \ldots - r_{ik}^2$ may be taken to indicate the lack of precision of the ith test as an instrument for measuring those factors with which we are concerned, the variance of the 'individual' or 'specific' factor for i being in this way incorporated into u_i^2. Now, with the traditional method of least squares, when the measuring instruments differ in their precision, the optimal estimates are found by minimizing the sum of the *weighted* squares of the residuals; and for this purpose the weights (usually denoted by h_i^2) will be the reciprocals of the residual variances, u_i^2. In carrying out the actual computations, the usual procedure consists in 'preparing' the observed values by multiplying each at the outset by the square root of the weight, i.e., by $h_i = 1/u_i$. The figures so obtained can then be treated as observed values of 'equal unit weight.'[2]

If we now revise the proof outlined above, in order to allow for this modification in our procedure, we shall reach a determinantal equation of the same general type as before, but $R - R_u = R_c$ will take the place of R. The final equation for the latent roots will consequently be (cf. this *Journal*, II, pp. 66, 113):

$$| R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}} - \lambda I | = | R_a - \lambda I | = 0, \qquad \text{(xvi)}$$

where $R_a \equiv R_u^{-\frac{1}{2}} R_c R_u^{-\frac{1}{2}}$ will denote the 'augmented' correlation matrix obtained

[1] It is tempting to suggest that the reliability coefficients might be computed from a single application by using the split-half procedure or (better still) the variance method (Burt, C., *Brit. J. Educ. Psychol.*, XV, p. 91). But in my experience this does not yield satisfactory results.

If analysis of variance is used when the same tests have been applied to the same persons on two occasions, it is easy to show that the variance due to the interaction between persons and trials (or occasions) is equal to the square of the error of measurement, i.e.,

$$I^2/(N-1) = s_i^2(1 - r_{ii'}),$$

where I denotes the interaction. This gives for the ordinary reliability coefficient $r_{ii'} = 1 - I^2/s_i^2(N-1)$. With m applications the denominator becomes $(m-1)(N-1)$.

[2] Cf. Johnson, W. W., 'The Theory of Errors and Method of Least Squares,' 1892, pp. 98f., or Aitken, A. C., *Statistical Mathematics*, 1939, p. 111.

by ' correcting ' the ' reduced ' correlation matrix (R_c) for ' specificity.' [1] A numerical solution of the equations will necessitate a process of successive approximation, since R_u can only be determined after we have reached some provisional decision as to how many common factors are significant and what are their respective variances.

If there are s significant common factors, the correlation matrix of rank s based on the procedure just described will yield a much closer fit to the observed correlation matrix (of rank n) than we should obtain with s principal components obtained by Pearson's procedure or with any of the modifications so far mentioned. This will be readily seen if we consider the case in which $s = 1$. For instance, in my 1909 investigation already cited, the first issue to be decided was the presence of other significant factors over and above the first or ' general ' factor. Spearman in correspondence had maintained that only one factor was operative in tests such as I had used for my table.[2] Now, suppose the table *had* formed a ' perfect hierarchy ' : then, if Pearson's procedure had been adopted, the first principal component, obtained by inserting units in the leading diagonal, would have given a decidedly poor fit to the observed coefficients, and a better fit could only have been obtained by introducing supplementary factors, which might justly have been criticized as artefacts. Hence it seemed desirable to substitute the ' highest common factor ' that could possibly be secured under the conditions specified. This entails inserting ' reduced ' variances (or self-correlations) in the leading diagonal instead of variances of unity.[3]

The residuals left after the significant factors have been deducted from the ' corrected ' correlations are virtually partial correlations. This will be clear if we translate the matrix equations into ordinary notation. Thus, if the hypothesis is that only *one* significant common factor (g) is present (i.e., that R_e is a ' perfect hierarchy '), the residuals are

$$r_{ij \cdot g} = \frac{r_{ij} - r_{ig}r_{jg}}{\sqrt{1 - r_{ig}^2}\,\sqrt{1 - r_{jg}^2}}. \qquad \text{(xvii)}$$

After k factors have been extracted, the residuals will be

$$r_{ij \cdot g \ldots k} = \frac{r_{ij} - r_{ig}r_{jg} - \ldots - r_{ik}r_{jk}}{\sqrt{1 - r_{ig}^2 - \ldots - r_{ik}^2}\,\sqrt{1 - r_{jg}^2 - \ldots - r_{jk}^2}} = \frac{\text{residual}}{\sqrt{(1 - i\text{th communality})}\,\sqrt{(1 - j\text{th communality})}} \qquad \text{(xviii)}$$

[1] The phrase is not a very happy one. In Germany, about the beginning of the present century, instead of terms of French origin, like *Präzision*, it had become customary to use words of German etymology, like *Zuverlässigkeit*. Writing in Germany, therefore, and taking up Cattell's suggestion that the ' precision ' of a test should be estimated by the correlation between two applications, Spearman and Krueger spoke of estimating its *Zuverlässigkeit*. On returning to this country, Spearman adopted the usual word to re-translate this term, namely, ' reliability ' ; and so ' reliability ' has come into general use among psychologists in place of the older term ' precision.' The word ' specificity ' was used by Spearman to cover all factors peculiar to each test ; and in his early papers anything that was not attributable to the one and only common factor was treated by him as part of the ' unreliability ' of the test. Hence he extended the formula derived by Yule to correct for attenuation due to ' unreliability ' (in the narrower sense of chance fluctuation) in order to correct for attenuation due to any form of ' unreliability.' This procedure magnified the importance of the *least* reliable of the tests. The formula described in the text was, therefore, put forward as providing a more satisfactory procedure for ' correcting for unreliability ' in the broader sense. However, to avoid confusion between the two formulæ for correction, the newer procedure was described as a method of ' correcting for *specificity* ' : (cf. *Brit. Ass. Ann. Rep.*, 1915, p. 695, and (5), Table XX.).

[2] An attempt to prove this was later given by him in his joint paper with Hart (' General Ability : Its Existence and Nature,' *Brit. J. Psych.*, V, 1912, pp. 54f.).

[3] It was not difficult to show that the important change was not the ' correction ' but the use of ' reduced ' test-variances. Indeed, in adapting Pearson's procedure for psychological work I ventured to introduce this change from the start by inserting ' reduced self-correlations ' for unity in the leading diagonal.

To test these 'corrected residuals' for their statistical significance various devices may be suggested.

(*a*) As is evident from eqn. (xvi), the factorial analysis is tantamount to determining the principal components of the 'corrected' correlation matrix. Accordingly, provided the investigator is willing to embark on the necessary iterations, the significance of the principal components could be determined by the chi-squared test for latent roots as described above.

(*b*) In psychological work, however, it is unusual for an investigator to undertake to calculate the factors by the lengthy iterative methods necessary for 'weighted summation.' As a rule he is content to apply 'simple summation' to the table of observed correlations, and to allow for 'specificity' by simply substituting 'reduced self-correlations' for unity in the leading diagonal. Formulæ (xvii) and (xviii) may still be applied ; and the results may still be treated as partial correlations. Hence we can, if we wish, test each 'corrected' residual separately by the ordinary standard error for a zero partial correlation, $1/\sqrt{N}$. This was in fact the test adopted in earlier investigations on educational abilities.[1]

(*c*) If a joint significance test is required for the whole aggregate of corrected residuals, two procedures may be suggested. First, we may add their squares, and so compute their variance or their standard deviation, and then compare the result with the value to be expected if they were due merely to chance. The calculation is still rather lengthy. To abridge it we can average the squares of the *uncorrected* residuals, and then divide the result by (1 − average communality). The final criterion (naturally somewhat rough and ready, but fairly safe for all but borderline cases) can be expressed as follows : [2]

$$\frac{\text{S.D. of residuals}}{1 - \text{av. communality}} \leqslant \frac{1}{N}. \qquad \text{(xix)}$$

(*d*) The defect of these last two procedures lies in the fact that they take into account only the number of persons. But, as the scheme set out above (Table I) reminds us, the significance of the aggregate of residuals also depends on the number of degrees of freedom supplied by the tests ; and this, we have seen, necessarily diminishes with the extraction of each fresh factor. Hence a more satisfactory procedure (particularly for borderline cases) would be to sum *the squares of the corrected residuals*. When N is large, this sum will be distributed approximately as chi-squared. For the reasons set out above, the degrees of freedom remaining at any given stage will be computed in accordance with eqn. (x). We can therefore determine the significance of any particular set of residuals by reference to the usual statistical tables. Although not as exact as the first (method *a*), this procedure yields much the same verdicts as regards significance. And we have thus arrived at a method which in some measure adjusts for the skewed distribution of correlation coefficients (see p. 118 above).

3. *Method of Maximum Likelihood applied to Correlations.* Dr. Lawley [3] has

[1] Burt, C., *Distribution and Relations of Educational Abilities*, 1917, pp. 54-7.

[2] This criterion has been independently suggested by McNemar, who arrives at a more elaborate formula by a rather circuitous algebraic proof, and then omits terms assumed to tend towards zero. He proposes the criterion for cases where the factors have been extracted by the centroid method ((18) pp. 9–18).

[3] Lawley, D. N., 'The Estimation of Factor Loadings by the Method of Maximum Likelihood,' *Proc. Roy. Soc. Edin.*, LX, 1940, pp. 64–82. In a paper submitted to the Royal Society Symposium in 1938 (in which I was much indebted for help to Professor R. A. Fisher, who was at that time my colleague at University College), and again in *Factors of the Mind* (1940, p. 259), I had briefly suggested the possibility of applying the method of maximum likelihood to factorial and other problems ; I understand that the same suggestion was also made about the same time by Dr. G. Brown of Princeton University. The more obvious lines of proof were discussed in *Laboratory Notes on Factor Analysis* (1934). I should point out, however, that the example quoted from those *Notes* by Mr. Dixon (this *Journal*, V, pp. 59–60) was deliberately selected because it magnifies the differences between the results obtained. Usually they are much smaller.

My own procedure started with the conception that the factor-measurements for the several persons may be regarded as weighted means, and the argument followed the lines of Hotelling's generalization of Student's ratio, which leads to a ratio of two determinants (*Ann. Math. Stats.*, II, 1931, pp. 360f.). To obtain a working method certain simplifications and approximations seemed inevitable. These led back once again to the modifications of Pearson's method described above ; and it seemed doubtful in this particular case whether anything was in the end gained by the more elaborate approach. Dr. Lawley, following the standpoint introduced by Spearman, would seem to be more interested in the factor-saturations for tests than the factor-measurements for persons. He therefore assumes that, in applying a sample of n tests to a sample of N persons, we are drawing sample measurements from the *same* n-dimensional distribution of correlated variables, not from N different distributions. But these differences in standpoint raise side-issues which I hope to take up later.

endeavoured to provide a more rigorous theoretical basis for current factorial procedures by starting with the principle of maximum likelihood. This principle he employs in order to obtain the most likely fit for the observed variances and covariances or correlations, not (as might be expected) for the observed measurements. His choice has provoked criticisms from Gale Young and several other writers ; but it finds a practical justification in the intractable equations to which the alternative method eventually leads.[1]

Lawley's procedure starts with the joint probability distribution of the sample variances and covariances as given by Wishart and Bartlett,[2] and ends with a measure of the discrepancy between observation and hypothesis in which the essential term is the ratio of the determinants of the observed and the hypothetical matrices of variances and covariances respectively. To avoid the calculation of the determinants, he suggests as an approximation taking the sum of the squares of what I have called ' corrected residuals ' with the degrees of freedom as given by eqn. (x).

Bartlett's derivation (25) leads to a slight modification of Lawley's main formula, namely,

$$\chi^2 = -N' \log_e \frac{|R|}{|R_0|}, \qquad \text{(xx)}$$

where R denotes the sample matrix of covariances and variances (or correlations) and R_0 the ' true ' matrix, and the multiplier N' should be, not N (the number of persons in the sample), but $N - \frac{1}{2}(2n + 5)$. If (he adds) it could be shown that, with the so-called method of maximum likelihood, we can derive ' efficient ' estimates [3] for R_0, then these could be substituted for R_0 in eqn. (xx). The number of degrees of freedom will be

$$\tfrac{1}{2}(n - k)(n - k - 1) - n. \qquad \text{(xxi)}$$

The chief objection to most of the methods described in this section is that they are even more lengthy and laborious than the preceding. From a theoretical standpoint the determinantal formula reached by Bartlett would seem undoubtedly the best, and his mode of approach appears to indicate the direction in which we must look for a more adequate treatment of the whole subject. But from a practical standpoint the requisite calculations are exceedingly tedious for the ordinary worker to carry out except when dealing with extremely small correlation tables. Lawley's contributions mark a valuable advance in the mathematical treatment of the subject. But the simplification he suggests involves two rather rough approximations, first for the determinants and then for the logarithm. Although the test is based nominally on the principle of maximum likelihood, the supplementary assumptions and approximations which are introduced would appear to prevent it from claiming the merits of any rigorous deduction from that principle. Since no clear distinction is made between the ' specific factors ' and the ' error factors,' the presence of the former still tends to magnify the postulated common factors ; and thus, as indeed practical trials suggest, the criterion proposed usually errs on the generous side, and so leads to the acceptance of more factors than are really warranted : the psychologist is then tempted to supply some over-ingenious interpretation. There is one further feature that to my mind makes the whole procedure rather dubious. If, owing

[1] Young, G., ' Maximum Likelihood Estimation and Factor Analysis,' *Psychometrika*, VI, 1941, pp. 49–53. In a later paper (*Proc. Roy. Soc. Edin.*, LXI, 1942, pp. 176–85), Lawley replies, and endeavours to reach a working method, based on the more obvious procedure—a method which he calls ' Method II.' The effort, however, is admittedly not very successful. As Bartlett points out ((24) p. 81), " the second method would at first seem the more fundamental, but is known to lead to difficulties." He adds that " M. G. Kendall has also raised doubts about the suitability of the maximum likelihood method in the present context." Bartlett himself briefly puts forward alternative suggestions which deserve further exploration (for a fuller discussion, see *J. Roy. Stat. Soc.*, B, XII, 1950, pp. 72, 86).

[2] Wishart, J., ' The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population,' *Biometrika*, A, XX, 1928, pp. 32–52 : cf. Wishart and Bartlett, *Proc. Camb. Phil. Soc.*, XXIX, 1933, pp. 260–70 ; and Kendall (19), pp. 330–5.

[3] Lawley assumes this without proof. But, as Bartlett has pointed out, in order to obtain a practicable method of computation, Lawley has introduced certain *ad hoc* assumptions : hence a formal proof seems really needed ((25) pp. 81–4).

to the size of the sample and the number of supplementary factors, some one test has a ' communality ' of nearly 1·00, that test then acquires an excessively high weighting : (if its communality was 1·00, it would receive infinite weight). At the same time the working formulæ for the factor saturations still do not give the *best* fit to the observed correlations that is obtainable with the number of factors presumed to be significant. For this and other reasons it still seems safest in most practical enquiries to adhere to the simple Pearsonian principle of seeking the ' closest fit.' If a more thorough investigation of significance were needed, it would in my view be better to spend the extra time and labour, not on these extremely protracted calculations, but on a re-application of the tests, so that an empirical basis is available for distinguishing error or ' unreliabilities ' from specific factors in the narrower sense : (thus reverting to eqn. (v) rather than eqn. (vii)).

IV. TESTS FOR CURRENT FACTORIAL PROCEDURES

The more elaborate methods of calculating factor-saturations which our proofs have for the most part hitherto assumed are chiefly of theoretical interest, and are seldom employed in actual practice. In the majority of factorial investigations the computer usually employs some approximate procedure, such as simple summation (eqn. (xi)) or the centroid method, and seldom requires a very precise test of statistical significance. His main problem is to know when his common factors are ceasing to possess any evidential value—in short, when to stop factorizing and when to cease interpreting. As a rule, therefore, some short approximate method will generally suffice.

1. *Testing Isolated Residuals.* In researches carried out with a properly constructed battery of tests on a sample of reasonable size, it is commonly found that, after the first few factors have been extracted, there is only a comparatively small number of residuals that are large enough to deserve consideration. At this stage the simplest procedure is generally the safest : test these larger residuals, as they stand, by the standard error of the *observed* correlations.[1] Such a test may have to be re-applied for two or more successive factors. When at last a borderline factor is reached, the *whole* set of residuals should be tested systematically in the same way, and the factor considered devoid of intrinsic significance, unless more than 1 in 20 of the residuals on which it is to be based exceed twice the standard error.[2]

[1] This was the method introduced in my 1909 paper, where the observed and theoretical correlations, the discrepancies between the two, and the relevant probable errors were systematically set out in a single table.

Holzinger's table of probable errors for correlation coefficients, obtained from samples of increasing size, will be found serviceable for this purpose (*Statistical Tables for Students in Education and Psychology*, 1928, Table X). Occasionally, instead of examining the largest *residuals*, it is wiser to examine the largest figures to be obtained from the factor that is intended to fit them, i.e., the products of the largest *saturations :* usually the two sets of figures will be of much the same size, but at times the products will be smaller. The general rule will be the same : unless more than 1 in 20 exceed twice the *s.e.* of the observed correlations, the factor can claim no significance whatever.

Two methods similar to the above have not infrequently been adopted in more recent investigations, namely, testing the residuals (i) by the standard error of a *zero* correlation or (ii) by the standard error for the difference between *two observed* coefficients. But a residual correlation is not itself an observed correlation ; it is a difference between an observed and an expected value ; and, since the expected value is fixed by the terms of the hypothesis to be tested (viz., so many factors with just these saturation coefficients), it does not possess a sampling error.

[2] It should, however, be remembered that significance is a relative term ; even though a factor may not be significant as judged by some arbitrary level, its presence may nevertheless be more probable than its absence. If the decision as to the suspected factor is of importance, it would be wiser to adopt some kind of sequential procedure by continuing the experiment until more conclusive evidence one way or the other is secured. In such a case, when deciding whether it is worth while extending the experiment, it will be helpful to consider the pattern formed by the larger residuals. If the few large residuals are clustered in a small group of tests, their presence is more suggestive than if they are scattered irregularly over the table ; on the other hand, if they are all confined to a single test, their presence is less suggestive, since the errors of correlations are themselves correlated.

2. *Testing the Aggregate of the Residuals.* As we have already seen, however, it is extremely convenient to have a single expression for the joint significance of *all* the residuals. With simple summation the residuals in the entire table necessarily sum to zero. Hence it is natural to treat the average of their squares as a variance, and its square root as a standard deviation : we may then compare this value with the standard deviation to be expected when the residuals are the result of mere chance. To determine this value several alternatives have been put forward.

(*a*) A common procedure is to compare the observed value with the standard error of a *zero* correlation, namely, $1/\sqrt{N}$. This can hardly be defended.[1]

(*b*) A slightly better procedure is to compare the figure with the standard error of a correlation equal to the *average* of the observed coefficients.[2]

(*c*) If each residual is first divided by the standard error of the observed coefficient, the standard deviations of these ratios should be approximately equal to unity when the residuals are due to chance.[3] An abridged calculation based on this principle consists in making the division after instead of before the process of averaging. Thus as a crude approximation we can divide the standard deviation of the residuals (taken as they stand) by the standard error of the mean observed coefficient : (this is virtually the method of the preceding paragraph carried one step further). Simpler still, to save the labour of squaring, we might use the mean deviation (instead of the standard deviation) in both cases.

From empirical studies it would appear that, with the doubtful exception of the third, these rough-and-ready methods of obtaining a joint significance-test for all the residuals are not very trustworthy. They are, however, still sometimes adopted, especially by American investigators, to check or supplement the results obtained with other criteria.

3. *Chi-Squared Test with z-Transformation.* With nearly all the significance-tests based directly on the coefficients of correlation, one of the main difficulties, as we have seen, springs from the fact that, unless the coefficients are low, their distribution is far from normal. Hence, as a general rule, the only satisfactory way of testing a discrepancy between two values for the same correlation will be to use Fisher's z-transformation ($z = \tanh^{-1}r$). Let us write for the expected value of an observed correlation

$$\bar{r}_{ij} = r_{ig}r_{jg} + r_{ih}r_{jh} + \ldots + r_{ik}r_{jk}, \qquad \text{(xxii)}$$

where $g, h, \ldots, k$ denote the k common factors which by hypothesis are presumed to be sufficient to explain the observed value, r_{ij}. If there was only one discrepancy to test, the regular procedure ((6) p. 189, ex. 31) would be to convert both r_{ij} and $\bar{r}_{ij}$ to the corresponding z-values, and test the significance of the difference between z_{ij} and $\bar{z}_{ij}$: (the effect of the transformation is to enlarge the difference $r_{ij} - \bar{r}_{ij}$ to much the same extent as ' correcting ' the residual by the partial correlation formula as described above). Now, the distribution of $z\sqrt{N-3}$ is nearly normal, with unit standard deviation. It follows that, on summing the squares for *all* the

[1] The objection is the same as that given in the previous footnote. The practice of comparing residuals with the standard error of a *zero* correlation has been commoner in educational enquiries than in psychological, and is now rather unusual. In a current article, however, J. G. Ingham applies it, together with two other criteria, to establish the presence of a memory factor. He notes that the method yields a different verdict from the others, which he apparently prefers (*Brit. J. Psych., Gen. Sect.*, XLIII, 1952, p. 26).

[2] Cf. Guilford, J. P., *Psychometrika*, p. 494, and references. Guilford himself " is inclined to believe that this criterion is too crude." In a more recent publication, he adopts a method similar to that suggested in footnote 1, p. 125, namely, calculating the product of the two highest factor-saturations, but he proposes to compare the result with the standard error of a *zero* correlation (' Printed Classification Tests,' *Army and Air Forces Psychological Research Report*, No. V, 1947).

[3] This procedure is put forward briefly and tentatively by Thurstone in the only sentence referring to significance tests in *The Vectors of Mind*, p. 149 : from a later publication ((9), pp. 66f.) it would appear that he considers Tucker's criterion the best. In *Multiple Factor Analysis* there seems to be no reference to the subject. Reyburn and Taylor (*Brit. J. Psych.*, XXX, 1939, pp. 151–65) prefer to divide each residual by the probable error and consider whether the frequency distribution of the ratios differs significantly from a normal curve with an *S.D.* of 1·48.

discrepancies and multiplying the result by $(N-3)$, we obtain a quantity which will be distributed approximately as chi-squared. Consequently, to test the significance of the aggregate of the residuals, we may take

$$\chi^2 = (N-3)\Sigma(z-\bar{z})^2, \qquad \text{(xxiii)}$$

with the same number of degrees of freedom as before (eqn. (x) above).[1] For this purpose it is quickest to use a table which enables the computer to enter with r and read off z, instead of vice versa as with Fisher's table : (such a table, compiled from figures for the hyperbolic tangent in the *Smithsonian Mathematical Tables*, pp. 86 *et seq.*, is given in the *Notes* already cited). In spite of obvious defects, both theoretical and practical, I still consider this to be the most useful of all the tests available when current factorial procedures are employed.

V. DIRECT TESTS OF FACTOR-SATURATIONS

For many subsidiary problems it would be particularly helpful to have a simple and trustworthy formula for estimating the standard error of a single factor-saturation. Often, for instance, an investigator may desire to know whether a given test is worth retaining, or whether the differences between the saturations furnished by two or more tests are statistically significant. With orthogonal factors, the factor-saturations specify the correlations between the factor and the tests. Hence, in seeking for a quick criterion, it is tempting to begin by considering the saturations as coefficients of correlation, and the factor-variances as squares of multiple correlations. On this basis one simple negative rule appears both safe and useful. If a coefficient would not be significant when treated as an ordinary observed correlation, *a fortiori* it will not be significant when occurring as a hypothetical saturation ; and the same with differences between two saturations. On the other hand, figures that pass this rough test cannot be accepted forthwith as significant, since the criterion is too generous.

1. To derive a formula for the standard error, we may start in the usual way by taking logarithmic differentials and then using the Pearson-Filon equations (1) for the correlation between correlation coefficients.[2] This leads to exceedingly complicated algebraic expressions. But, by introducing plausible assumptions and neglecting terms involving higher powers, some kind of approximate formula can eventually be obtained. None seems to be altogether satisfactory : they are either too simple to be exact or too elaborate to be readily applied.

(*a*) The following[3] illustrate the type of formulæ so reached : they vary according to the nature of the assumptions introduced to simplify the expressions.

$$\sigma_r = \frac{1}{\sqrt{N}}\left\{\frac{1}{2}\left(1-2r^2\right)\left(1+\frac{1}{\bar{r}}-r^2\right)+\frac{1}{4}\left(1+\frac{1}{\bar{r}^4}\right)\right\}, \qquad \text{(xxiv)}$$

where r denotes the saturation coefficient to be tested and $\bar{r}$ the mean of the saturation coefficients for the factor in question. When there is no great difference between the saturations for the same factor, and r does not differ greatly for $\bar{r}$, then (very approximately)

$$\sigma_r = \frac{1-r^2}{\sqrt{N}}\sqrt{\left\{1+\frac{3}{4\bar{r}^2}\right\}}. \qquad \text{(xxv)}$$

Practical trials suggest that the values so obtained are rather too high in the case of the saturations obtained for the first or second factor, but fairly satisfactory in the case of later factors, for which it is more likely to be needed.

[1] Cf. Burt (15), p. 339. As there stated, the method has a fairly long history of trials in a wide variety of research theses, and has been found on the whole to give satisfactory results. A similar significance-test based on the z-transformation is occasionally used in non-factorial studies where an aggregate test for a group of independently obtained correlations is wanted : cf., for instance, Tippett, L. H. C., *Methods of Statistics*, 1931, p. 143.

[2] For an example of this procedure, see Kelley, T. L., *Crossroads in the Mind of Man*, 1927, pp. 40f.

[3] Burt, C., *Mental and Scholastic Tests*, p. 139 ; cf. *Ann. Eugen.*, XIII, p. 251.

(*b*) However, the foregoing, like many of the other criteria proposed, assumes that (i) the number of persons (N) is alone likely to affect the standard error to any appreciable degree. Now, as we have seen, there can be little doubt that (ii) the number of tests correlated (n), and (iii) the number of factors already extracted (k) should also be taken into account in judging statistical significance. On making further allowance for these, we reach the following rough approximations : [1]

$$\sigma_r = \frac{1 - r^2}{\sqrt{N}} \sqrt{\left\{1 + \frac{k+1}{n-(k+1)}\right\}}$$

$$= (1 - r^2) \sqrt{\frac{n}{(n-k-1)N}} . \qquad \text{(xxvi)}$$

The numerical values can readily be obtained by using a table of standard or probable errors for ordinary correlations, and multiplying the tabulated figure by the ratio introduced into the equation to allow for the number of tests and the number of factors extracted. This procedure has given fairly satisfactory results when tried empirically with data for which there is independent evidence in regard to significance or distribution.

2. One or two writers have derived special formulæ for the factor-saturations that are reached by their own particular methods of factorization. For example, Lawley has developed a formula for the variance of a factor-saturation, when obtained by his procedure, which he expresses as follows :

$$\sigma_{ia} = \frac{\theta_\alpha}{N}\left\{1 - \sum_{j=1}^{j=a} \theta_j f_{ij}^2 + \frac{1}{2}\theta_a l_{ia}^2\right\} , \qquad \text{(xxvii)}$$

where N is the number in the sample, f_{ij} the factor-saturation for test i in factor j, $\theta_\alpha = 1 + \frac{1}{\nu_a^2}$, and ν_a^2 is the variance for a, the factor in question, and l_{ia} is the saturation of test i with factor a.[2]

Starting with Spearman's formula for a squared saturation based on a ' triad ' of three observed correlations, and Kelley's derivation of its standard error (*loc. cit.*, p. 41), Holzinger and Harman have also deduced several alternative and approximate formulæ for testing the significance of a saturation coefficient obtained by their bifactor method. As applied to factor analyses carried out by simple or weighted summation, their formula for the standard error of a saturation coefficient would seem to give values that are usually too small, and that for the standard error of the corresponding residual would seem to give values that are decidedly too large ; as a consequence, with one and the same set of data the formulæ occasionally yield somewhat conflicting results. But, as the authors themselves observe, their " sampling error formulas . . . do not apply for other types of factor-saturations."[3]

VI. MISCELLANEOUS METHODS

Whether concerned with single saturations or with saturations treated in the aggregate, most of the significance-tests considered so far have been based on much the same broad principles as those adopted for other statistical constants, namely, on the comparison of an observed deviation (or an observed standard deviation) with a hypothetical standard deviation, or, more conveniently still, on the ratio of an observed variance to a hypothetical variance. However, many writers have produced ingenious expedients of their own, derived from original principles holding good for factorial problems only. As we shall see, very few enable the investigator to estimate the odds for or against the hypothesis that is under examination. Hence, at best they can serve merely as supplementary guides. The most plausible depend on certain features which a matrix of residuals may be expected to exhibit when it is the result of accidental fluctuations.

1. If, after k factors have been extracted, the residuals are due mainly to chance, their positive and negative signs will be scattered at random. In that case, no rearrangement of the order of the variables will succeed in eliminating all the larger negative residuals from the positive quadrants or all the larger positive residuals from the negative quadrants ; and consequently no reversal of signs

[1] Burt, C., and Banks, C., *Ann. Eugen.*, XIII, p. 255.
[2] Lawley, D. N., ' Problems in Factor Analysis,' *Proc. Roy. Soc. Edin.*, A, LXII, 1949, pp. 1–6 ; also this *Journal*, II, p. 97, and refs. Lawley has also obtained formulæ for the standard errors of residuals. He has, however, since found that the estimated saturations are correlated ; hence the standard error obtained by the formula given in the text is correct " only if the values of saturations for the previous factors are regarded as already given " (this *Journal*, III, 1950, p. 76).
[3] Holzinger, K. J., and Harman, H. H., *Factor Analysis*, 1941, pp. 124-32.

in the columns or rows will convert the entire table of residuals, or even the larger residuals in it, to positive form. Actuated by this idea, Coombs (17) has calculated the number of irreducible negatives to be anticipated in batteries of different sizes. With a battery containing less than 10 tests, 33·3 per cent. of the residuals may be expected to remain negative at the critical stage; with larger batteries the proportion steadily increases, and reaches 43 per cent. with 50 tests. Such a criterion, however, scarcely seems precise enough to be relied on as conclusive; and, like most of the criteria of this type, it makes no reference to the size of the sample.

2. Ledyard Tucker [1] has derived the following empirical criterion for 'completeness of factorization':

$$\sqrt{\left\{\frac{\sum |d|_{k+1}}{\sum |d|_{k}}\right\}} = \frac{n-1}{n+1}, \qquad \text{(xxviii)}$$

where $|d|$ denotes the sum of the absolute residuals (i.e., of their numerical values regardless of sign) after k or $k+1$ factors have been extracted, each sum including the residuals appearing in the diagonal as well as in the rest of the matrix. As a rough survey of the literature shows, this criterion has been employed, especially by American factorists, more frequently than any other. It is in keeping with the familiar fact that, when we attempt to factorize a large bipolar matrix of correlations due solely to chance, we obtain successive sets of residuals whose average absolute value (or whose average square-sum) diminishes hardly at all as one factor after another is extracted.[2] Thurstone, who regards it as the most effective of the devices he mentions, observes that "as yet there is not available any rational proof for this criterion"; but he apparently believes that it will turn out to be "an approximation of a rational criterion."

Four criticisms can be made. (i) It is difficult to accept a test of significance which appears to make no allowance for the number of tests or the number of factors extracted, and above all includes no explicit reference to the size of the sample. (ii) An attempt at 'rational proof' leads one to expect the radical to appear on the right-hand side of the equation, not on the left; alternatively, if the right-hand side is to be a ratio involving the number of tests (not the square root), then one would expect on the left-hand side a ratio of sums of squares, not of simple sums. Indeed, provided the ratio on the right-hand side is based, not on the number of tests but on the number of degrees of freedom remaining, a rational derivation along these lines will readily produce a fairly good criterion similar in form to eqn. (xxviii), but in that case the radical on the left-hand side disappears ((15) p. 349). (iii) As it stands, Tucker's formula proves to be far too generous, particularly when (following Thurstone's method of factorization) the highest residual in each column is taken as the residual communality: thus, with a set of tetrachoric correlations obtained from 218 to 240 persons only, Thurstone concludes that as many as twelve factors are significant ((9) p. 70). This seems scarcely credible with so small a sample. (iv) When tried with samples where the dispersion or the error is known, the formula is found to give erroneous verdicts ((11) p. 17).

3. Starting from a somewhat similar principle, and arguing from the dispersion of latent roots, Gale Young (12) has derived an 'index of clustering':

$$\zeta \equiv \frac{n\sum\sum r_{ij}^2 - S^2}{(n-1)S^2} \leqslant \frac{n-k}{(n-1)k}, \qquad \text{(xxix)}$$

where r_{ij} denotes an element of the product-sum or correlation matrix R, and S is the sum of the latent roots (i.e., with weighted summation, the sum of the diagonal elements in R). The 'null value' is derived by showing that, when the matrix R has $(n-k)$ zero roots, then its vectors will be arranged randomly in the k-space if ζ has the value given on the right of eqn. (xxix). Once again, however, it is difficult to accept a criterion which makes no reference to the number in the sample.

4. Hoel (13) has deduced two tentative criteria, starting with the method of principal axes, but modifying the formulæ to suit methods based on simple summation and reduced communalities:

(*a*)
$$w = \sqrt{\frac{N}{2n}\left\{\frac{|R|}{\varphi} - 1\right\}}, \qquad \text{(xxx)}$$

where φ denotes the sum of a series of products in which certain communalities are multiplied by their complements: e.g., for 5 tests and the first factor

$$\varphi = \Sigma(1 - H_1^2)\ldots(1 - H_{a4}^2)H_{a5}^2 + (1 - H_1^2)\ldots(1 - H_5^2).$$

When $w \geqslant 1{\cdot}65$, the odds are 19 to 1 against the truth of the hypothesis that the factor under consideration is significant.

(*b*)
$$K_q + H_m^2 \leqslant \frac{1}{1 - 1{\cdot}65\sqrt{2/N}}, \qquad \text{(xxxi)}$$

where K_q is the qth latent root of the correlation matrix with self-correlations of unity, and H_m^2 the smallest of the reduced self-correlations.

[1] I give the revised form of Tucker's criterion (*Psychometrika*, V, 1940, p. 125).

[2] For the way factor-variances usually tend to flatten out, see *Factors of the Mind*, Table VII, p. 358. Their coefficients of variation may not do so particularly when no allowance is made for specifics.

Since the first criterion entails the calculation of the determinant of the correlation matrix, it is scarcely a method for general use. The second requires the calculation to be carried only as far as the factors are likely to be significant. In the few trials I have made, the pronouncements of the two criteria are not always consistent. Hoel himself appears to prefer the first method, as being " the more sensitive."

5. Several other methods of testing significance deserve a passing reference, though none seems now to be in frequent use. Kelley extended Spearman's tetrad-difference procedure by introducing a pentad criterion (the usual determinantal criterion for a matrix of rank 2). More recently he has derived a ' variance ratio test.'[1] Hotelling briefly discusses the number of significant factors present when his method of determining principal axes is employed.[2] Wilson and Worcester (10) have described a chi-squared test based on the expression for the quadratic form which occurs in the exponent of e in the formula for a normal multivariate distribution ; the requisite computations, however, become quite intractable when the number of tests is large. Swineford suggests continuing factorization until the distribution of the residuals ceases to be significantly skew : for a more rigorous test she proposes calculating the correlation between the observed correlations and the corresponding residuals, and continuing factorization until this correlation ceases to be significant.[3]

Other tentative proposals (nearly all since discarded) are summarized in an article by Mosier (11). He himself suggests (from a single empirical trial) that the best criterion is " the maximum distance of the centroid from the origin $\left(\sum_{j}^{n} \sum_{k}^{n} r_{jk},\ j \neq k \right)$ obtainable by reflection of tests." But this conclusion appears partly to result from the nature of the trial matrix he employed. He agrees with McNemar that Thurstone and his followers usually attempt to demonstrate far more factors than the size of the sample permits ; and his final conclusion is that, of the various criteria available at the date of his investigation, " none is wholly satisfactory."

VII. EMPIRICAL STUDIES OF SIGNIFICANCE-TESTS

Since the commoner methods of factor analysis are based on procedures which involve very considerable simplification, and sometimes on devices introduced for practical convenience with no analytic justification whatever, it seems desirable that anyone who wishes to defend a practical test of significance should first check its applicability by a series of empirical trials, applying it, for example, to subdivided samples drawn from one large group or to artificially constructed tables where the amount of error is predetermined. It was on the basis of trials such as these that my own formulæ (xii), (xxiii), and (xxvi) were ultimately recommended.

Of the few investigations carried out along similar lines, the most instructive is that of McNemar (16). He concludes that " the sampling behaviour of first centroid factor loadings is much like that of ordinary correlations, whereas that of other loadings is disturbingly large " ; in general the empirical dispersions for the second factor prove to be about double that of the first, and those for the third factor about three times as large. He further finds that the distributions tend to be skewed (much as we might expect if high factor loadings behaved like high correlations), and that for factors after the first they tend towards bimodality. On the whole, therefore, McNemar's findings would support the use of some such formula for single saturations as that proposed above (eqn. (xxvi), p. 128).

In our own empirical studies we find that, with the first factor, the standard deviations for the saturations are nearly always a little larger than those for observed correlations of the same size, and that, as more and more factors are extracted, the dispersions steadily increase. With our data the rate of increase is perhaps a little

[1] Kelley, T. L., *Crossroads in the Mind of Man*, 1927, pp. 58f., ' A Variance Ratio Test of the Uniqueness of Principal Axis Components,' *Psychometrika*, IX, 1944, pp. 199f. (The latter is similar to the usual *F*-test, but the degrees of freedom take into account the number of rotations required to transform the correlation matrix to its principal axes.)

[2] Hotelling, H., ' Analysis of a Complex of Statistical Variables into Principal Components,' *J. Educ. Psychol.*, XXIV, 1933, pp. 437–41.

[3] Swineford, F., ' Comparisons of Multifactor and Bifactor Methods of Analysis,' *Psychometrika*, VI, 1941, pp. 375–82.

slower at first, and a little more rapid later on, than the American investigations would suggest. This may conceivably be due to the fact that, in centroid analyses, the highest correlations (or the highest residual correlations) are usually inserted in the diagonal. Bimodality is liable to appear when the computer has used a battery consisting of sharply distinguishable groups of tests ; but frequently there may be no bimodality whatever.

Judging from investigations of this kind, it would seem that, broadly speaking, with batteries of between 10 to 20 tests of the usual type, and with a summational method of analysis,[1] at least 20 persons are needed to establish 1 factor, 50 to establish 2, 100 to establish 3, and between 200–400 to establish 4. I believe, as a rule, it is unwise to try to establish *more* than 4 factors (1 general and 3 bipolar) with any single battery or in any single research.

VIII. SUMMARY AND CONCLUSIONS

1. From the foregoing review, it is clear that none of the numerous significance tests that have been put forward is completely immune from objection on either theoretical or practical grounds or both. The more rigorous tests are all too elaborate for routine computations ; the more practicable rest on dubious simplifications. The theoretical difficulties, as we have seen, arise chiefly from two sources : first, the errors of the observed inter-correlations are themselves correlated, and secondly, the assumption of specific factors brings a further complication with which multivariate analysis appears still unable to deal. The practical difficulties are due partly to the appearance of large determinants in nearly all the basic formulæ, and partly to the fact that only a lengthy iterative procedure is capable of yielding factor-saturations and factor-measurements which are themselves free from the minor inaccuracies resulting from the usual mode of calculation. No doubt in the near future both these difficulties will be removed by further research or better computing appliances. Meanwhile, the theoretical comparisons we have made, together with an increasing experience of the practical merits of the different methods, would seem to point to certain provisional conclusions that may be helpful to research-workers who are contemplating enquiries of the customary type.

2. In all such investigations, as will now be evident, two problems must be distinguished.

I. *When to stop extracting factors.* During the last thirty years, it has become a recognized routine to start any factorial study with a preliminary analysis by the method of simple summation or some equivalent procedure, e.g., the centroid method. At this stage the obvious question to ask is : how many factors should be extracted ? For such a purpose, I suggest, a rough examination of the residuals, in the light of the relevant standard errors, will ordinarily suffice. But it is essential, particularly if the investigator intends to proceed afterwards to a group factor analysis or to a rotation of factors, to adopt a generous standard. Instead of demanding the usual odds of 19 to 1 before accepting the more questionable factors,

[1] With such a summational analysis, all the factors except the first will be bipolar. The corresponding number of group factors will be much larger : e.g., 3 summational factors (of which 2 would be bipolar) will in general yield 4 group factors of the narrower type as well as one basic factor, and (if subdivided factors are used) 2 broad group factors as well, making 5 to 7 in all.

The requirements suggested in *Factors of the Mind* (p. 357) were perhaps a little too stringent : they referred to a period when comparatively small batteries were in use. However, McNemar would prefer a still more severe criterion. His statement (*loc. cit.*, p. 148) that increasing the number of ' variables ' (i.e., tests) reduces the number of significant factors seems to me somewhat dubious. On that and other grounds he proceeds to " question the significance of more than *two* factors for samples of size 200 " (p. 145). However, considering the readiness with which some investigators will extract ten or a dozen factors from samples containing less than 200, his warning is undoubtedly salutary.

any factor whose presence seems decidedly more probable than not should in general be retained. If a more definite rule is required, I would recommend (*a*) that with small samples the computer should extract at least one factor more than those that are fully significant, and (*b*) that with large samples he should adopt a critical ratio of $1 \times s.e.$ instead of $2 \times s.e.$ Moreover, it will occasionally be found that the extraction of one or two supplementary factors which are well below the level of significance may be requisite to prevent the common factor-variances exceeding 1·00.

II. *When to accept factors as established.* When the whole analysis has been completed, and the investigator is seeking to establish and interpret some new factor, the position is different. Then a more austere standard should be observed. At the same time, the discovery that, with the sample used, a given factor turns out to be non-significant cannot be taken to prove (as Spearman so often assumed) that it is non-existent. Unless some kind of sequential scheme can be adopted, probably in most researches the best principle to follow would be to distinguish between those factors which appear unquestionably significant (i.e., beyond the 0·05 level) and those for which the evidence is suggestive though not conclusive (say, between the levels of 0·25 and 0·05).

At this stage I suggest the following procedures, differing according to the method of factor analysis employed.

(A) With weighted summation :

(i) When the Pearson method of factorization has been employed (extracting principal components from a correlation table with units in the diagonal), there can be little doubt that Bartlett's test is by far the best.

(ii) When allowance has to be made for specific factors, a full investigation would, in my view, include ' correction for specificity,' and then, (*a*) if all the latent roots have been calculated, the investigator may apply Bartlett's test in its original form (eqn. (xiii), p. 119) or, (*b*) if only a few of the latent roots are to be calculated, his modification of Lawley's chi-squared formula (eqn. (xx), p. 124). However, unless special facilities for mechanical computation are available, the extra labour needed will rarely be practicable.

(iii) When weighted summation is used with reduced self-correlations, but without correction for specificity, then the chi-squared test applied to the z-differences seems to provide the most convenient test (eqn. (xxiii), p. 127).

(B) With simple summation and group factor analysis no very rigorous procedures are available. I suggest that—

(i) As the constants for each factor are finally computed, the investigator should note (*a*) whether the successive factor-variances are beginning to flatten out ; (*b*) whether there are numerous negative residuals which cannot be abolished by sign changes ; (*c*) whether there is still a marked clustering among the residuals (when suitably rearranged) ; and, as these indications suggest that the significance of the remaining factors is becoming increasingly doubtful, he should (*d*) carry out a systematic test of the residuals by comparing each with twice the standard error of the observed correlation (p. 125) ;

(ii) For a final overall test of the significance of the most doubtful factor, the investigator may again use the chi-squared test based on the z-differences (eqn. (xxiii), p. 127) ;

(iii) To determine the significance of a single saturation (or of the difference between two saturations) he may for most purposes rely on eqn. (xxvi) (p. 128) ;

(iv) If factor-measurements are to be calculated with simplified weights, he would do well to attempt an analysis of variance (p. 116).

3. The defect of all the foregoing methods is that they oversimplify the analysis of ' error.' Thus, the influence of errors arising from the sampling of tests and the sampling of the occasions is either wholly ignored or else merged with that of errors arising from the sampling of persons : at the same time sampling errors generally are left indistinguishable from other kinds of irrelevant influence, notably that of the so-called ' specific ' or ' individual factors.' These are points that seem to call for some revision both of practical and of theoretical procedures. So far as actual practice is concerned, the limitations can in my view only be overcome by adopting a more elaborate experimental scheme, with a view to reducing specific

factors to a minimum or converting them to group factors, and securing objective data for estimating unreliability. In any thoroughgoing factorial investigation it would seem best to revert to the old experimental plan in which all tests are applied to the same group more than once. With two applications it becomes possible to determine the reliability of the several factors or to apply the alternative methods outlined by Bartlett ((25) p. 80, and refs.). But, wherever possible, more than two applications should be attempted, since in that way alone can we study the presence of trends and interactions among the several factors. For these purposes the analysis of variance (in the simpler cases), or (in the more complex cases) the somewhat elaborate methods of multivariate analysis developed in econometrics, may be profitably applied. On the theoretical side, as Bartlett has already observed, the most pressing requirement is " a detailed re-examination and formulation of the most appropriate statistical procedure for dealing with the usual structural equation," i.e., the equation which includes specific factors as well as common factors of the more familiar type.

REFERENCES

1. Pearson, K., and Filon, L. N. G. (1898). ' On the influence of random selection on variation and correlation.' *Phil. Trans. Roy. Soc.*, CXCI, A, 229–311.
2. Pearson, K. (1901). ' On lines and planes of closest fit to systems of points in space.' *Phil. Mag.*, 6th Ser., II, 559–72.
3. Burt, C. (1909). ' Experimental tests of general intelligence.' *Brit. J. Psych.*, III, 94–177.
4. Yule, G. U. (1910). *Introduction to the Theory of Statistics.* (2nd ed., Yule, G. U., and Kendall, M. G., 1937.) London : Griffin and Co.
5. Burt, C. (1917). *The Distribution and Relations of Educational Abilities.* London : P. S. King and Son.
6. Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh : Oliver and Boyd.
7. Pearson, K., and Moul, M. (1927). ' The mathematics of intelligence : I. The sampling errors in the theory of a generalized factor.' *Biometrika*, XIX, 246–91.
8. Thurstone, L. L. (1933). *The Vectors of Mind.* Chicago : University of Chicago Press.
9. Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago : University of Chicago Press.
10. Wilson, E. B., and Worcester, J. (1939). ' A note on factor analysis.' *Psychometrika*, IV, 133–48.
11. Mosier, C. I. (1939). ' Influence of chance error on simple structure.' *Psychometrika*, IV, 33–44.
12. Young, G. (1939). ' Factor analysis and the index of clustering.' *Psychometrika*, IV, 201–8.
13. Hoel, P. G. (1939). ' A significance test for minimum rank in factor analysis.' *Psychometrika*, IV, 245–54.
14. Hsu, P. L. (1939). ' On the distribution of the roots of certain determinantal equations.' *Ann. Eugen.*, IX, 250–8.
15. Burt, C. (1940). *Factors of the Mind.* London : University of London Press.
16. McNemar, Q. (1941). ' On the sampling errors of factor loadings.' *Psychometrika*, VI, 141–52.
17. Coombs, C. H. (1941). ' Criterion for significant common factor variance.' *Psychometrika*, VI, 267–72.
18. McNemar, Q. (1942). ' On the number of factors.' *Psychometrika*, VII, 9–18.
19. Kendall, M. G. (1943). *The Advanced Theory of Statistics.* London : Griffin and Co.
20. Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton : Princeton University Press.
21. Wilks, S. S. (1947). *Mathematical Statistics.* Princeton : Princeton University Press.
22. Burt, C. (1947). ' A comparison of factor analysis and analysis of variance.' *Brit. J. Psych., Stat. Sect.*, I, 3–26.
23. Saunders, D. R. (1948). ' Factor analysis : some effects of chance error.' *Psychometrika*, XIII, 251–8.
24. Bartlett, M. S. (1948). ' Internal and external factor analysis.' *Brit. J. Psych., Stat. Sect.*, I, 73–81.
25. Bartlett, M. S. (1950). ' Tests of significance in factor analysis.' *Brit. J. Psych., Stat. Sect.*, III, 77–85.
26. Burt, C. (1950). ' The influence of differential weighting.' *Brit. J. Psych., Stat. Sect.*, III, 105–25.
27. Kendall, M. G. (1950). ' Factor analysis as a statistical technique.' *J. Roy. Stat. Soc.*, B, XII, 60–94.
28. Kendall, M. G. (1951). ' Regression, structure, and functional relationship.' *Biometrika*, XXXVIII, 11–25.
29. Bartlett, M. S. (1951). ' The effect of standardization on a chi-squared approximation in factor analysis.' *Biometrika*, XXXVIII, 337–44.

BOOK REVIEWS

Fundamental Statistics in Psychology and Education. By J. P. Guilford. 2nd Edition. New York : McGraw Hill Book Co. Pp. xiv + 633. £2 2*s.* 6*d.*

Professor Guilford's earlier work on *Psychometric Methods* is familiar to almost every student of psychology in this country. Some years after it appeared, Professor Guilford felt a growing need for supplementing it by a text-book specifically devoted to statistical procedures. Published during the war, the first edition of this newer book never became well known to readers in this country. But the revised edition should form an excellent introduction for all who wish to understand or use the devices of modern statistics either in psychology or in education.

The volume has been largely rewritten. Far more space is now allotted to analysis of variance, the reliability and validity of mental tests, and different procedures for scaling. Methods appropriate to small samples are more fully discussed ; and special attention is devoted to the logic underlying statistical inferences. One of the most interesting features arises from the fact that " factor theory is now regarded as essential for the understanding of such phenomena as inter-correlation, multiple prediction, and correction for attenuation " ; and, in accordance with this conviction, Professor Guilford expounds the problems of validity and reliability very largely in terms of constituent factors. This brings his standpoint close to that already adopted by statistical psychologists in this country who " followed the early lead of Galton and Karl Pearson."

The exposition throughout is clear and well arranged. For the benefit of those whose mathematical training is limited, many of the arguments which are commonly expressed in algebraic form are also given verbally in the shape of logical deductions ; and the beginner is greatly assisted by the free use of concrete illustrations and diagrams. Each chapter ends with a series of exercises for the student ; but many readers will perhaps miss the usual lists of detailed references for each particular topic.

Unlike most other books dealing nominally with the same general subject, this is by no means a mere account of standard statistical procedures with mathematical proofs. It demonstrates in the concrete how such procedures are helpful in dealing with the various practical problems that arise in psychology and education, and emphasizes both the advantages and the limitations of each method. The treatment is at once eclectic and impartial ; and the whole forms a valuable text-book for senior courses.

R. F. Douglas

Theory of Mental Tests. By Harold Gulliksen. London : Chapman and Hall, Ltd. Pp. xx + 486. $6.00.

Professor Gulliksen rightly insists on the need of " a firm grounding in test theory as a basis for any work in test development." Accordingly, he has attempted to " bring together in one readily available source the chief technical developments in test theory during the last fifty years." Primarily the book is concerned with the *statistical* aspects of the ' theory of mental tests.' The methods of planning and constructing such tests, of choosing or phrasing the test-problems, or of selecting and arranging the content, and the advantages and disadvantages of the different procedures, fall outside his purview.

After a brief introductory chapter, the book opens by setting out a series of ' basic equations ' derived from the definitions of random error and a true score. It then deals in turn with errors of measurement, of substitution, and of prediction, and their interpretation. This leads to a study of the effects of test length on means, variances, reliability, and validity. A chapter describing statistical criteria for parallel tests presents comparatively new procedures, due chiefly to Wilks and Votaw. The problem of group heterogeneity is then discussed at some length, with special reference to its effects on coefficients of reliability and validity ; and this is followed by an account of methods for correcting the decrease in heterogeneity resulting from selection. After briefer discussions of familiar topics such as the experimental methods of obtaining test reliability and the various questions raised by speed tests and power tests, the book concludes with a series of chapters on methods of scoring, standardizing and equating scores, differential prediction and weighting, and item analysis. Factor analysis and analysis of variance are not separately discussed, partly, no doubt, because systematic presentations of these techniques are already available in volumes appropriated to those subjects. Otherwise, as will be seen, nearly all the standard topics and procedures are dealt with, and fitted into a logically evolving sequence, which presents a systematized context. Professor Gulliksen also includes several valuable contributions of his own, notably on the theory of items, of incidental selection in multivariate data, and of ' power ' tests that are partly speeded.

As the long and useful bibliography shows, Professor Gulliksen has relied mainly on American sources. Full recognition is accorded to the work of English statisticians, like Yule, Pearson, and Fisher. But, apart from references to the earlier contributions of Spearman and later publications of Burt, the British student will learn little of the more specific contributions made to the theory of mental testing by psychologists in his own country.

The mathematical treatment is at once simple and rigorous. Except for two or three chapters, which assume a knowledge of maxima and minima in the calculus and the use of determinantal methods to solve simultaneous linear equations, or (in certain sections only) some acquaintance with the elementary theory of determinants and matrices, the proofs are so arranged as to require only the more elementary algebraic or geometrical procedures included in nearly all school courses.

The derivations of the various formulæ are prefaced by clear and lucid discussions of each problem. Assumptions and definitions are carefully formulated ; and, owing largely to the use of different styles of type, the proofs, the conclusions, and the summaries of the several chapters are set out with admirable clarity. Numerous figures, diagrams, and nomographs are included to illustrate the various methods or to facilitate practical calculations. Each chapter concludes with a convenient recapitulation, and a set of problems and exercises. The appendices contain useful tables and lists of equations, and two series of examination questions, the first an examination in statistics to be used as a review at the beginning of a course on test theory, and the second an examination on test theory itself, to be given at the end.

The book will form an indispensable handbook for all who may be engaged in the work of constructing, validating, and standardizing tests ; and is unquestionably the best existing text-book on its subject for use in senior University courses.

Cyril Burt

This Matter of Mind. By Bryan H. Kirman, M.D., D.P.M. London : C. A. Watts & Co. Pp. 92. 1*s*.

A popular book on psychology may seem scarcely to require notice in this journal. Dr. Kirman, however, devotes a number of pages to describing and criticizing what he takes to be the methods and the views of statistical psychologists ; and, since he himself is a psychiatrist of standing, and has published researches of his own in various medical journals, his statements, if left without contradiction, may be widely accepted as authoritative and accurate. Such pronouncements are bound to increase the popular prejudice against statistical surveys and quantitative research, and so make the work of the social or educational psychologist even harder than it is.

One of his main aims is to " expose the fallacies underlying so-called intelligence tests," and to deprecate the preoccupation with the mathematical approach, which, he declares, " has alarmed the public with gloomy prophecies." " The position," he says, " is still dominated and overshadowed by the highly mathematical and speculative work of Pearson : this worker, using results obtained in an effort to measure various human abilities, deduced that all these had factors in common ; this common factor he calls ' *g*.' " " His followers, dizzy with their enthusiasm for mathematical calculations, go much farther. . . . Often it is assumed that the data are free from errors of sampling." And " on page after page covered with algebraic formulæ they demonstrate that ' *g* ' has an independent reality : ' *g* ' in fact is a little god who lies concealed in a man's brain and determines his fate," and the fate of nations too.

On enquiring what are the sources for these various statements, we learn that the outstanding instances are to be found in the account of intelligence as given in *The Young Delinquent* and in " various broadcast talks." " As an example of the fallacies of abstract reasoning in consideration of a practical problem," Dr. Kirman proposes to take " Burt's argument " demonstrating " a calamitous fall in the national intelligence." Dr. Kirman devotes five pages to explaining and criticizing what he states are " the main steps in the development of this argument." First, he says, Burt begins by assuming that intelligence, as measured by tests, is inherited ; next he " assumes that people of lower intelligence tend to be more fertile " ; thirdly, combining " his two generalizations," he deduces by mathematical calculation the imminence of a " calamitous fall."

A reference to the " broadcast talks " as printed in the *Listener* discovers nothing of this kind. Presumably, therefore, Dr. Kirman is quoting the published memorandum prepared by Burt at the request of the Royal Commission on Population. Its object, however, was not to show that any imminent fall or decline had already been established, but merely to make out a case for a systematic investigation of the problem. The statements that differences in intelligence as measured by tests are largely inheritable, and that people of lower intelligence (as thus measured) tend to be more fertile, are not advanced as mere *assumptions :* various lines of evidence are offered. But Burt then goes on to insist that mere mathematical deductions developed on this basis are of themselves bound to be inconclusive. Instead of armchair statistical deductions from a single survey, he proposes a comparison of successive surveys at widely separated intervals. What information of this type we possess already does *not*, he says, confirm the sombre prophecies of the more pessimistic followers of Pearson. Anything like a " ' galloping plunge towards intellectual bankruptcy ' (as one writer has described it)

would," he points out, " have inevitably aroused spontaneous comments from school teachers and from educational officials." As nothing of the kind has been noted, he himself puts forward several suggestions explaining how the expected effects of the birthrate might in fact have been counteracted by " unsuspected biological mechanisms." Since his memorandum was published, investigations of the kind proposed have in fact been commenced, though, as the investigators themselves insist, it is still much too early for any clear conclusion to be established.

Dr. Kirman's repeated references to " Pearson's notion of '*g*' as a factor common to all abilities " once more shows that he is quite content to criticize various authorities without troubling to ascertain what they actually said. Pearson himself never used psychological tests, and was in fact almost as sceptical of their value as Dr. Kirman. So far from introducing the idea of " a common factor called '*g*,' " he was one of its earliest and most vigorous opponents. The " notion of *g* " is of course due to Spearman, not Pearson. Dr. Kirman's own method of refuting any such hypothesis consists in citing two cases in which an individual committed to an institution, on the ground that he is mentally defective in certain respects, happens to show high mental abilities in certain other respects : the idea of using a standardized test to measure the degree of deficiency or of ability, or of ascertaining the frequency of such cases, he evidently regards as at once pedantic and fallacious.

In most of his other references to the work of statistical writers Dr. Kirman manages almost uniformly to attribute to the persons he names views which are the exact opposite of what they really stated. In *The Young Delinquent*, for example, there is no reference to '*g*' ; and nothing about the " independent reality " of the common factor. So far from containing " page after page covered with algebraic formulæ," it gives scarcely a single formula. The later volume on *The Factors of the Mind* certainly includes a section on the " reality of factors " ; but here it is argued that factors like '*g*' are merely abstract hypothetical concepts with no more " real existence " (and no less) than the concept of energy in physics.

The conception of the underlying logic of statistical psychology put forward in this book is shared by many other psychiatric writers. They have heard that mathematical arguments are " essentially deductive," by which they understand deducing particular conclusions or concepts from general propositions or concepts : they therefore suppose that statistical psychologists, " like the metaphysicians of old," are trying to *deduce* the present existence of this or that ' real entity ' or the future occurrence of this or that particular event. Such an attempt would admittedly be (in Dr. Kirman's phrase) " a glaring example of the fallacies of abstract reasoning in the consideration of practical problems." It therefore seems desirable to point out, as explicitly and emphatically as possible, that the arguments as set forth by the factorists quoted are essentially inductive,[1] not deductive : the use of algebraic formulæ or arithmetical computation is merely incidental. In short, as has often been emphasized in the pages of this *Journal*, the statistical parts of such inferences represent nothing but a firm resolve to use, wherever possible, rigorous and exact methods of verification in place of the rough and ready methods of jumping to conclusions that apparently content Dr. Kirman and those who think with him.

E. M. Moursy

[1] By ' inductive ' reasoning is meant, not, as is so often supposed, deriving a general proposition from the examination of a number of particular instances, but the verification of one out of two or more mutually exclusive hypotheses by disproving all the alternatives except that one : see, for example, Fisher, R. A., *The Design of Experiments*, ch. II, or (if Dr. Kirman insists that he is attacking not Fisher's statistical procedures, but Burt's) *The Factors of the Mind*, pp. 37f.

Vol V
Part III

The British Journal of Psychology,
Statistical Section

November,
1952

Dr. WILLIAM BROWN

The unexpected death of Dr. William Brown is a great loss to British psychology. He was one of its most versatile representatives. His numerous contributions to psychiatry and his work for the British Psychological Society will doubtless be described in other journals. Here I should like to pay a tribute to his work as a pioneer in the statistical field, and to record my own indebtedness to his friendly help and criticisms.

Brown was the first to expound and organize, for use in psychological research, the newer statistical methods worked out by Pearson and his school. At Oxford, where he was a scholar of Christ Church, he was my senior by a year ; and the diversity of his achievements had already created something of a legend. At an unusually early age he had gained first-class honours in Physics at London. He came to Oxford in 1899 ; took Mathematical Moderations, Final Honours in Natural Science, and then, by an unusual feat, *Literæ Humaniores* as well. At the close of this period he joined the tiny band who, at McDougall's instigation, were beginning to explore possibilities in the new field of mental tests. Spearman, who had just returned from Wundt's laboratory, settled down for a short time nearby, and carried out sensory tests in a local village school. Keatinge, the new Reader in Education, started testing memory on the lines of Ebbinghaus. McDougall himself devised new tests of attention and apperception for school children, and listened gravely to Brown's advocacy of partial regression equations for combining the results.

After winning the John Locke Scholarship in Mental Philosophy, Brown returned to London to attend Pearson's courses at University College. In 1907 Spearman became Sully's assistant at University College, and a year later Brown became head of the Psychological Department at the rival institution of King's College. With Pearson's assistance he then planned a research which should meet the most rigorous statistical requirements. Tests were to be applied to far larger groups of children and students than Flugel and I had been able to secure at Oxford ; and almost all Pearson's technical devices were systematically worked out. His final conclusions confirmed the view that " tests fall into a number of subgroups," but failed to confirm the notion of a " central factor " common to all. The results were incorporated in a thesis on ' The Use of Correlation in Psychology,' for which he was awarded, in 1910, both the D.Sc. and the Carpenter Medal. The verdict of both psychologists and statisticians was expressed by Rivers : " If statistical research is at all possible in psychology, here for the first time is a model of what such a research should be."

In the following year he published the first edition of his book on *Mental Measurement*. Its object was to make " the biometric methods of Professor Karl Pearson and his school more generally known among psychologists." Slender as it was, it included many original suggestions, and more than one new formula which is still associated with his name. The first part dealt with the psychophysical methods as a means of mental measurement, and gave a simplified account of Pearson's system of frequency-curves. The second part, an expansion of his doctoral thesis, dealt with correlational techniques, and was " intentionally controversial." Pearson's novel terminology, the technical nature of his proofs, the laborious computations which his formulæ entailed, and his uncompromising insistence on the need for large samples had combined to deter most psychological research-students from even thinking of adopting his methods ; and Pearson himself never applied his ' method of principal axes ' to more than four variables. Spearman's simplified procedures furnished a welcome alternative. The ' foot-rule ' for calculating correlations, the contention that small samples with ' correction for unreliability ' were better than large samples with hasty experimental techniques, above all, the claim that each test contained only one common factor over and above the specific factor peculiar to itself made it possible for psychological students and teachers to break new ground by attacking simpler problems first and by using very simple methods. The divergent results so obtained, however, seemed to demand a more rigorous approach ; and Brown accordingly thought it essential to subject the new devices that had become so popular to a more rigorous examination.

What was needed, so he maintained, was a simplification of Pearson's exposition, not of Pearson's techniques ; and hence much of the second part of the book was devoted to

examining Spearman's formulæ and criticizing his main research. With my own contributions he dealt more leniently, since I had kept closer to Pearson's method and had recognized the existence of supplementary factors in addition to general intelligence. Nevertheless, I too had departed from the full rigour of Pearson's requirements : the evidence for special abilities had been based (as it still usually is) on the simple residuals " instead of applying the full formula for partial correlation," and in each of my Oxford experiments the number of children tested was " far too small." On both these counts I came largely to agree with him ; and the next set of researches (planned with Moore in 1911) owed much to his published criticisms and still more to his friendly discussions and suggestions.

During the next two or three years the meetings of the little British Psychological Society were constantly enlivened by debates on the issues thus raised. At Oxford, at Cambridge, and at London, Brown and Spearman appeared side by side. Before an audience, at once bewildered and entertained, they plunged into amiable but animated polemics on the merits and defects of Pearsonian methods, each snatching the chalk from the hand of the other to dash down figures and formulæ on the blackboard. Even if few followed the mathematical arguments, an eager interest in quantitative procedures was successfully aroused, not only among the junior students who were privileged to attend, but also in the minds of older sceptics like McDougall, Myers, and Rivers. The excitement continued session after session until at length a still more heated controversy was set going by the novel theories of Freud ; and here too Brown took a leading and a lively share.

Meanwhile he had taken further degrees in medicine and surgery ; and during the First World War served as Neurologist to the Fourth Army in France. For a time he was Resident Medical Officer at the Maudsley Hospital, and directed experimental research in clinical psychology at Bethlem. Two further editions of *Mental Measurement*, greatly enlarged, were produced in collaboration with Godfrey Thomson. These, " like the first, were opposed to the theories of Professor Spearman " ; but Brown hoped " later on to add a Part III, summarizing results and theories in non-mathematical language, since at present the battle rages so hotly that this is not desirable, even were it possible." In 1921 he succeeded McDougall as Reader in Mental Philosophy at Oxford, and became Director of the new Institute of Experimental Psychology which he succeeded in establishing " in that stronghold of philosophical learning." He thereupon set to work, with the aid of William Stephenson, to carry out a still more elaborate research on the lines suggested " by Professor Pearson in a critical article on The Theory of Two Factors." This entailed the calculation of 14,535 tetrad differences. He still found evidence for the influence of group factors, such as verbal ability ; but, when the influence of these was removed, the results, as he candidly admitted, now seemed to confirm the presence and importance of the common factor, *g*.

In 1930 he was elected a Fellow of the Royal College of Physicians ; and his work as a consultant in London and a lecturer in Oxford, together with the preparation of books and papers on psychiatric subjects, prevented him from continuing his statistical investigations. After the Second World War was over, he lent invaluable assistance in supporting the foundation of this *Journal*, and was always willing to give advice on editorial problems. A few months before his death he wrote to me that he now at last felt free to resume his own interest in statistics. He was eager to produce yet another revision of *Mental Measurement*. In the very first version he had included a chapter explaining Pearson's development of the " general theory of multiple correlation," and the formula for the " correlation ellipsoid," which played an important part in the reduction of the multivariate distribution to terms of the principal axes. " Feeling dissatisfied with my attempt at a simple and intelligible account " (he wrote), " I omitted the chapter in the later editions ; but I now hope to revive it and expand it and explain its bearing on methods of factorial analysis. Pearson to-day seems almost forgotten ; yet I fully agree that his approach was not only the first but also the best."

During the last year of his life he was President of the British Psychological Society, and was an active host during the Oxford meeting earlier in the present year. He then seemed to have many years of vigorous life before him. In his letters and conversation he was full of his new projects. He had collected a vast amount of literature on factorial methods ; and was busily engaged on the work when he so suddenly died. Among British psychologists his mathematical knowledge and skill were unique. Had he been able to resume his studies in the statistical field, he might once again have become the leading authority on the subject in this country.

C. B.

Vol. V
Part III

The British Journal of Psychology,
Statistical Section

November,
1952

THE ESTIMATION OF A POINT BISERIAL COEFFICIENT OF CORRELATION FROM A PHI COEFFICIENT[1]

By WILLIAM B. MICHAEL

The Testing Bureau, The University of Southern California, Los Angeles, California

NORMAN C. PERRY

Department of Psychology and Statistics, San Jose State College, San Jose, California

and J. P. GUILFORD

Department of Psychology, The University of Southern California, Los Angeles, California

I. *Introduction.* II. *The Problem.* III. *Derivation of Formula for Estimation of Point Biserial Coefficient from Knowledge of a Phi Coefficient.* IV. *Determination of the Amount of Systematic Error in Formula* (6). V. *Calculation of Theoretical Values of the Phi and Point Biserial Coefficients.* VI. *Illustrative Example.* VII. *Relative Degree of Accuracy of Estimates Furnished by Formula* (6). VIII. *Interrelationships Among Parameters.* IX. *Summary.*

I. INTRODUCTION

In the statistical evaluation of the degree of internal consistency and validity of aptitude- and achievement-test items the coefficient of correlation usually considered to be logically most defensible is the point biserial. The use of a correlational approach is actually a problem of prediction in which the continuous dependent variable (usually represented by performance measures on an external criterion or by total test scores) is a function of several independent variables (items) of a dichotomous ('pass' or 'fail') nature. Typically the intended correct response to an item is thought to constitute one category of behaviour which is essentially qualitatively different from a second category made up of various responses keyed as incorrect. As a rule scoring weights of + 1 and 0 are assigned respectively to the 'right' response and to the one or more 'wrong' responses. It is evident that the typical 'point' nature of the independent (item-response) variable and the usual continuity of the dependent variable (test scores or criterion scores) constitute properties that are consistent with the assumptions appropriate to the use of the point biserial coefficient of correlation.

Unfortunately, the calculation of a point biserial coefficient of correlation is a time-consuming task. With respect to either of the two conventional formulas employed the mean of the scores of the group responding in a designated manner (e.g., 'pass' or 'fail') must be computed *for each item.* (In one of the formulas a mean is required for each of the groups 'passing' and 'failing' the item.) In addition, the standard deviation of measures on the continuous dependent variable is required.

[1] The contents of this article, one of a series of publications concerning the relationship of the phi coefficient to other coefficients, are based on a paper read before a session of the Western Psychological Association during its annual meeting in Fresno, California, 1952.

Experience has shown that in item-analysis work the most quickly calculated or estimated correlational statistic is the phi coefficient. Prepared by Guilford (1, 2) for use with extreme groups of equivalent size selected from an entire criterion sample, an abac permits a rapid estimation of a phi coefficient when the proportions of individuals in the ' upper ' and in the ' lower ' extreme groups responding in a designated manner to an item are known.[1] It is to be shown that a point biserial coefficient may be quickly approximated from knowledge of the value of a phi coefficient which, it may be assumed, has been conveniently estimated from Guilford's abac. In addition several patterns of interrelationships among the point biserial, phi, and tetrachoric coefficients, which are a function of levels of item difficulty and of proportions of individuals of the total criterion group found in the extreme groups, will become apparent in the entries of Table I and Table II.

II. THE PROBLEM

Specifically, it is the purpose of the writers

(1) To develop a formula that will permit the estimation of a point biserial coefficient from a phi coefficient when the proportion of individuals p passing an item in the two extreme groups combined is known and when the proportion p' which each of the extreme groups represents of the total criterion sample is given ;

(2) to describe a means for estimating the amount of systematic error present in values of the point biserial coefficient furnished by the formula ; and

(3) to present a set of tables based on the use of percentages of 50 and approximately 27 in the contrasted groups, which for various levels of item difficulty will

(*a*) present the (theoretically determined) magnitudes of point biserial coefficients corresponding to various phi coefficients associated with specified values of the tetrachoric coefficient,

(*b*) furnish corresponding formula estimates of the point biserial coefficient, and

(*c*) yield percentages of systematic error arising from use of the formula.

With respect to all three of these objectives the central approach is to relate both the phi coefficient and point biserial coefficient to the tetrachoric coefficient. It is assumed (1) that the dependent variable representing the criterion, or total test, scores is continuous and normal in form and (2) that the regression of the criterion variable upon the independent variable constituting item-response behaviour is linear. In the process of relating the phi coefficient and point biserial coefficient to the tetrachoric coefficient, it is necessary to make a third assumption—namely, that the independent variable is normally distributed in the instance of the tetrachoric coefficient. (The dichotomization of this normally distributed variable in terms of those passing or failing an item and the assignment of weights of + 1 and 0 to the proportions of area above and below the corresponding point of dichotomy on the unit-normal function taken to describe item-response behaviour constitute the necessary operations for the preservation of the point nature of item reaction when a point biserial or phi coefficient is computed.)

[1] Henceforth it will be assumed for convenience that a high score denotes above average performance. Therefore the upper extreme group is defined as one with relatively high total test or criterion scores, and the lower extreme group is the one with relatively low scores. The quotation marks will not be subsequently employed to set off the words *upper* and *lower*. For convenience a designated manner of response will be defined as " passing " or responding " correctly " to an item. Quotation marks will not be used subsequently about such words as *pass* or *fail*.

The fulfilment of these three assumptions (which is prerequisite to the realization of the amount of systematic error in the formula and to the calculation of various entries within the two tables) permits (1) a description of the functional relation between the dependent and independent variables in terms of the normal correlation surface ; (2) the use of a set of tables giving *Volumes of Normal Bivariate Surface*, prepared by Pearson (8), which serve as an indispensable aid in the calculation of theoretically expected values of the phi coefficient associated with specified magnitudes of the tetrachoric coefficient and with selected levels of item difficulty ; and (3) the development of a simple equation that expresses the point biserial coefficient as a linear function of the tetrachoric coefficient for any designated level of item difficulty. The theoretically expected values of the point biserial coefficient obtained from this equation can be compared with those approximated by the formula which furnishes an estimate of the point biserial coefficient from knowledge of the size of a phi coefficient. From the discrepancies observed the relative amount of systematic error may be readily calculated.

For future reference the following symbols may be defined :

Let x = continuous dependent, or criterion, variable (e.g., total test scores) for which a unit-normal distribution is assumed to exist in a generalized fourfold frequency table and upon which centile points are chosen for creation of extreme groups ;

y = the independent variable representing the behaviour of individuals to items, assumed to be a point distribution in the instance of the use of a point biserial or phi coefficient, but a (continuous) unit-normal distribution in the instance of a tetrachoric coefficient ;

p' = the proportion of individuals in each of the two extreme groups ;

h = the standard score designate of the x variable, numerically equal positive and negative values, $+h_0$ and $-h_0$, of which correspond to the centile points $100(1 - p')$ and $100p'$ respectively ($h_0 = 0$, when $p' = \cdot 50$; $h_0 \neq 0$, when $p' < \cdot 50$) ;

$y_{p'}$ = the two equal ordinates in the unit-normal distribution of the x variable found at the two centile points $100(1 - p')$ and $100p'$, separating the upper and lower extreme groups from a central group of individuals whose scores near the mean of the x variable are not contained among those of individuals in the contrasted groups ;

P = the level of item difficulty represented by the proportion of individuals in the total criterion sample who pass an item ($P = p$ when $p' = \cdot 50$) ;

$Q = 1 - P$, the proportion of individuals in the total criterion sample who respond incorrectly ;

p_U = the proportion of individuals in the upper extreme group who pass an item ;

p_L = the proportion of individuals in the lower extreme group who pass an item ;

$p = (p_U + p_L)/2$, the proportion of individuals in the two extreme groups combined who pass an item ;

$q = 1 - p$, the proportion of individuals in the two extreme groups combined who respond incorrectly ;

y_P = the ordinate in a unit-normal frequency distribution which separates the proportions P and Q ;

y_p = the ordinate in a unit-normal frequency distribution which separates the proportions p and q ;

k = the standard score designate of level of item difficulty (when $k > 0$, $P < \cdot 50$, and $Q > \cdot 50$; when $k < 0$, $P > \cdot 50$, and $Q < \cdot 50$) ;

r_t = the coefficient of tetrachoric correlation corresponding to given proportions within the cells of a fourfold frequency table ;

$\varphi_{p'}$ = the phi coefficient calculated from use of the upper and lower extreme groups each consisting of the proportion of individuals p' of the total criterion sample ;

r_{pb} = the point biserial coefficient of correlation existing for a total criterion sample ;

$\varphi_{p'}r_{pb}$ = the estimated point biserial coefficient of correlation furnished by formula (6) in which $\varphi_{p'}$ is the independent variable.

III. DERIVATION OF FORMULA FOR ESTIMATION OF POINT BISERIAL COEFFICIENT FROM KNOWLEDGE OF A PHI COEFFICIENT

In a recent article Guilford and Perry (3) developed one formula for estimating a tetrachoric coefficient of correlation from knowledge of a phi coefficient and a second formula for approximating a point biserial coefficient of correlation. Although taking into account various combinations of marginal proportions of the x and y variables in a fourfold frequency table, the formula was not applicable to the situation in which there was an omission of cases in the vicinity of the mean of either variable, as occurs in the instance of item-analysis procedures based on the use of contrasted groups each consisting of equal proportions p' less than ·50.

When correction was made in the phi coefficient for coarse grouping in both the x and y variables, the following estimate r_φ of the coefficient of tetrachoric correlation was obtained :

$$r_\varphi = \frac{\sqrt{p'q'}}{y_{p'}} \frac{\sqrt{PQ}}{y_P} \varphi, \tag{1}$$

where φ represents the phi coefficient calculated from a fourfold frequency table in which the sum of each set of marginal proportions ($p' + q'$ or $P + Q$) is equal to unity. In this formula p' and q' are not equal in value except when the size of p' and hence that of q' is ·50.

When only one of the variables of the fourfold frequency table is essentially continuous, say x, the correction for coarse grouping proposed by Guilford and Perry yields a formula which gives an estimate φr_{pb} of the point biserial coefficient :

$$\varphi r_{pb} = \frac{\sqrt{p'q'}}{y_{p'}} \varphi. \tag{2}$$

If $p' = q' = ·50$, (2) reduces to

$$\varphi r_{pb} = \frac{p'}{y_{p'}} \varphi, \text{ or,} \tag{2'}$$

since $y_{p'} = ·3989$, to

$$\varphi r_{pb} = 1·253\, \varphi. \tag{2''}$$

Perry and Michael (10) have shown that an estimate $r_{\varphi_{p'}}$ of a tetrachoric coefficient from knowledge of a phi coefficient $\varphi_{p'}$ may be obtained from the following formula :

$$r_{\varphi_{p'}} = \frac{p'}{y_{p'}} \frac{\sqrt{pq}}{y_p} \varphi_{p'}, \tag{3}$$

which representing a modification of (1) may be employed in conjunction with extreme groups of various, but equal, proportions. It is apparent that when $p' = q' = ·50$ formula (1) reduces to (3) and that when equal proportions exist in the two tails of the normal frequency distribution of scores on the x variable a more general expression for the factor used in correction for coarse grouping is $p'/y_{p'}$.

If it is assumed that $r_t \doteq r_{\varphi_{t'}}$ and that $r_b \doteq r_t$, it is apparent that since $r_{pb} = (y_P/\sqrt{PQ})r_b$ substitution of the right-hand expression of formula (3) as an approximation to r_b will yield

$$r_{pb} \doteq \left\{ \frac{y_P}{\sqrt{PQ}} \cdot \frac{p'}{y_{p'}} \cdot \frac{\sqrt{pq}}{y_p} \varphi_{p'} \right\}, \tag{4}$$

or, if it is assumed that the product $\frac{y_P}{\sqrt{PQ}} \cdot \frac{\sqrt{pq}}{y_p}$ is approximately unity,

$$r_{pb} \doteq \frac{p'}{y_{t'}} \varphi_{p'}. \tag{5}$$

Insertion of the previously defined symbol $\varphi_{p'} r_{tb}$ (which represents an estimate of r_{pb}) permits (5) to be expressed as follows :

$$\varphi_{p'} r_{pb} = \frac{p'}{y_{p'}} \varphi_{p'}. \tag{6}$$

It is evident that, as in formula (3), the expression required for correction of coarseness in grouping in the x variable is $p'/y_{p'}$. Since the point nature of the y variable is preserved by definition in the calculation of point biserial coefficient of correlation, only one correction factor is required, as is the case in formula (2). It is, of course, obvious that (2″) is but a corollary to (6).

Since in much practical work the proportions p_U and p_L are known it may be advantageous in some instances to employ a formula in which calculation of the phi coefficient is not required. Substitution of the familiar expression [1] $(p_U - p_L)/2\sqrt{pq}$ for $\varphi_{p'}$ in formula (6) yields upon simplification

$$\varphi_{p'}r_{pb} = \frac{p'}{y_{p'}} \cdot \left(\frac{p_U - p_L}{2\sqrt{pq}}\right), \tag{7}$$

which in the case that $p' = \cdot 50$ reduces to

$$\varphi_{p'}r_{pb} = \frac{\cdot 6267(p_U - p_L)}{\sqrt{PQ}}. \tag{7'}$$

In all the formulas presented a certain degree of systematic error exists. The right-hand expressions of (1) and (3) are actually equivalent to the constant term of the familiar infinite series function associated with the evaluation of a coefficient of tetrachoric correlation. Hence for values of r_t in excess of ·60 and for magnitudes of P greater than approximately ·75 or less than about ·25 the amount of systematic error is likely to exceed 5 per cent. (3, 7). Tables have been presented elsewhere, the entries of which describe the relative amount of error to be expected in formula (3) for a wide range of item difficulties ($k \leqslant | 1{\cdot}7 |$), for various proportions in the extreme groups (i.e., proportions corresponding to values of 0, ·50, ·60, ·70, and ·80 in h), and for tetrachoric coefficients varying between ·10 and ·80 (7).

Since the right-hand member of (3) was used as an approximation to r_b in the derivation of (5) and (6) it is evident that in (6) a degree of error exists for any fixed P that is proportional to the contribution of the items in the infinite-series function with powers higher than the first. Fortunately it is possible to determine for various combinations of h (or p'), k (or P), and r_t the absolute and relative amount of systematic error in $\varphi_{p'}r_{pb}$ through recourse to the tables compiled by Pearson (8).

IV. DETERMINATION OF THE AMOUNT OF SYSTEMATIC ERROR IN FORMULA (6)

Within Pearson's set of tables called *Volumes of Normal Bivariate Surface* the entries constitute numerous possible amounts of volume (i.e., proportions within the cells of a fourfold table) required to yield a designated magnitude of r_t for various combinations of dichotomies in the x and y variables. For the normal bivariate surface of unit volume the points of dichotomy in the x and y variables are denoted by h and k, which serve as the standard score designates of the marginal proportions in a fourfold frequency table.

Within each table (for which r_t is specified) entries are available for various combinations of magnitudes in h and k which are expressed in tenths of one unit. Each entry constitutes the proportion of volume standing on one of the quadrants formed by the intersection of the axes $x = h$ and $y = k$. Through use of two of Pearson's tables (one of which contains entries corresponding to a positive r_t ; the other, to a negative r_t) it is possible to ascertain for an r_t of given size the amounts of volume standing on two adjacent quadrants and hence the amounts (i.e., proportions in two adjacent cells of a fourfold table) of volume resting on the remaining two quadrants, since the magnitudes of the two sets of marginal proportions of the fourfold frequency table are fixed (in light of designated values in h and k).

For a given set of values in r_t, h (or p'), and k (or P) it is relatively simple to calculate from Pearson's tables the corresponding theoretical magnitudes of the phi coefficient $\varphi_{p'}$ and the point biserial coefficient r_{pb}. Once $\varphi_{p'}$ is determined, the estimate $\varphi_{p'}r_{pb}$ of the point biserial coefficient may be obtained from formula (6). The difference $\varphi_{p'}r_{pb} - r_{pb}$ constitutes

[1] The expression $(p_U - p_L)/2\sqrt{pq}$, employed whenever the proportions of individuals in the two extreme groups are equal, is equivalent to formula (8).

the (algebraic) amount of systematic error of this formula estimate, which may be defined simply as δ. The corresponding proportion of systematic error is given, of course, by $(\varphi_{p'}r_{pb} - r_{pb}) \div r_{pb}$, which may be designated as ε. Obviously, division of $\varphi_{p'}r_{pb}$ by $1 + \varepsilon$ yields a corrected value which within rounding is equivalent to r_{pb}.

V. CALCULATION OF THEORETICAL VALUES OF THE PHI AND POINT BISERIAL COEFFICIENTS

The calculation of the true $\varphi_{p'}$ from Pearson's tables which has been described in considerable detail in an article by Hertzka, Michael, and Perry (4) is effected through use of the formula

$$\varphi_{p'} = \frac{a - b}{n\sqrt{pq}}, \tag{8}$$

where

a = the proportion of volume standing on the first quadrant formed by the axes $x = h_o$ and $y = k$ (i.e., $a = p_U p'$, the proportion which passing individuals in the upper group constitute of the *total* criterion sample) ;

b = the proportion of volume standing on the second quadrant formed by a *different* set of axes $x = -h_o$ and $y = k$, except when $h_o = 0$ (i.e., $b = p_L p'$, the proportion which passing individuals in the lower group constitute of the total criterion sample) ;

$n = 2p'$, the proportion which the individuals in the two extreme groups combined represent of the total criterion sample, and

$p = (a + b)/n$, which is equivalent to the previous definition $(p_U + p_L)/2$.

The values for a and b are readily found in Pearson's tables, and the constants for the $\sqrt{pq}$ are easily evaluated from a set of tables prepared by Kelley (5).

The steps required for the determination of the theoretical value of r_{pb} are apparent from the two formulas expressed in standard score form that are conventionally employed in the calculation of a point biserial coefficient :

$$r_{pb} = (M_P - M_Q)\sqrt{PQ}, \text{ or} \tag{9}$$

$$r_{pb} = M_P \sqrt{\frac{P}{Q}}, \tag{10}$$

in which

M_P = the mean value, or centroid, with respect to the x variable of that portion of the volume of the entire normal correlation surface standing on the xy plane above the line $y = k$ (i.e., the mean score of the proportion of individuals in the *total* criterion sample who respond correctly to an item) ;

M_Q = the mean value, or centroid, with respect to the x variable of that portion of the entire normal correlation surface standing on the xy plane below the line $y = k$ (i.e., the mean score of the proportion of individuals in the total criterion sample who respond incorrectly to an item).

In mathematical terms, since the mean of a mass of unit density is equal to the sum of the first moments divided by the mass, it is apparent that

$$M_P = \left\{ \int_{-\infty}^{+\infty} x \left[\int_{k}^{\infty} f(x,y)dy \right] dx \right\} \Big/ P \tag{11}$$

and

$$M_Q = \left\{ \int_{-\infty}^{+\infty} x \left[\int_{k}^{-\infty} f(x,y)dy \right] dx \right\} \Big/ Q, \tag{12}$$

in which $f(x,y)$ is the familiar joint probability function describing the normal correlation surface.

It may be shown through rather involved integration, which is omitted, that

$$M_P = y_P r_t / P \tag{13}$$

and that

$$M_Q = -\, y_P r_t / Q. \tag{14}$$

Substitution of (13) and (14) into (9) yields

$$r_{pb} = \left(\frac{y_P}{P} + \frac{y_P}{Q} \right) \sqrt{PQ}\, r_t, \tag{15}$$

or, upon simplification, merely

$$r_{pb} = \frac{y_P}{\sqrt{PQ}}\, r_t, \tag{16}$$

which may be readily evaluated for given values in P (or k) and r_t.[1] From eqn. (8) and (16) the entries in the columns headed by $\varphi_{p'}$ and r_{pb} in Table I were computed relative to the corresponding values given for h, k, and r_t.

VI. ILLUSTRATIVE EXAMPLE

It may perhaps be helpful to illustrate the numerical operations involved in the calculation of a set of values in r_{pb}, $\varphi_{p'}$, $\varphi_{t'} r_{pb}$, and 100ε. Let it be assumed that $r_t = \cdot 50$, $k = \cdot 30$, and $h = \cdot 60$. Corresponding to $k = \cdot 30$ is a value of $\cdot 382088$ for P and $\cdot 381388$ for y_P. For $h = \cdot 60$ the magnitude of p' is $\cdot 274253$; and that of $y_{p'}$ $\cdot 333225^{-}$.

For a given level of item difficulty the determination of r_{pb} depends upon the evaluation of the constant $y_P / \sqrt{PQ}$ which is subsequently multiplied by the appropriate value of r_t. Hence by (16)

$$r_{pb} = \frac{\cdot 38139}{\sqrt{(\cdot 38209)(\cdot 61791)}}\,(\cdot 50), \text{ or } \cdot 39246^{-},$$

which is rounded to $\cdot 392$ in Table I.

The evaluation of the parameters a and b, which are found in Pearson's tables to be $\cdot 173370$ and $\cdot 043040$ respectively, is prerequisite to the calculation of $\varphi_{p'}$. Since $n = 2p' = \cdot 548506$, division of the sum $a + b$ by n, which is equivalent to $(p_U + p_L)/2$, permits the determination of p and hence of q which are to be substituted in the denominator of formula (7). It is found that $p = \cdot 216410 / \cdot 548506$, or $\cdot 39454$. Substitution of the values for a and b along with those of p and q yields

$$\varphi_{p'} = \frac{\cdot 173370 - \cdot 043040}{2(\cdot 27425^{+})\sqrt{(\cdot 39454)(\cdot 60546)}}, \text{ or } \cdot 48615,$$

which is rounded to $\cdot 486$ in Table I.

[1] It is apparent that the point biserial coefficient bears the same relationship to the tetrachoric coefficient as it does to the biserial coefficient. It is also evident that substitution of (13) and (14) into the expression for the biserial coefficient, $r_b = (M_P - M_Q)(PQ)/y_P$, will yield upon simplification $r_b = r_t$. The expression for r_b represents the slope of the straight line (describing the regression of x on y) which joins the centroids of the two portions of volume P and Q standing on the xy plane above and below the line $y = k$. (In a three dimensional representation a plane perpendicular to the xy plane and containing the line $y = k$ actually serves to separate the two portions of volume.) The co-ordinate points in the x and y variables defining the centroids of the volumes P and Q are respectively $(y_P r_t / P,\ y_P / P)$ and $(-\, y_P r_t / Q,\ -\, y_P / Q)$. The slope m of the regression line is the tangent of the angle of inclination of the line with respect to the y-axis (not the x-axis), since a score on the continuous variable x, e.g., a criterion, is being predicted from membership in one of two categories of the artificially dichotomized y variable (e.g., item response behaviour). Defining the slope m, the quantity

$$\left\{ \frac{y_P}{P}\, r_t - \left(\frac{-\, y_P}{Q}\, r_t \right) \right\} \Big/ \left\{ \frac{y_P}{P} - \left(\frac{-\, y_P}{Q} \right) \right\},$$

which is equivalent to $(M_P - M_Q)/y_P/(PQ)$, the expression for r_b, reduces to r_t. Hence $r_t = r_b = m$.

By formula (6) it is readily apparent that

$$\varphi_{p'}r_{pb} = \frac{(\cdot 27425^{+})}{\cdot 33322}(\cdot 48615^{+}), \text{ or } \cdot 40012,$$

which is entered as ·400 in Table I.

The absolute amount of systematic error δ is equal to the difference between $\cdot 40012^{-}$ and $\cdot 39246^{-}$, which is evaluated to be ·00766. Division of this difference by the value for r_{pb} yields the proportion of systematic error, which is equal to ·00766/·39246, or about ·02. This value is entered as a percentage of 2·0 in Table I.

In Table II only the percentages of systematic error are recorded. However, for each entry it was necessary to follow the procedure which has been outlined in this illustrative example.

VII. RELATIVE DEGREE OF ACCURACY OF ESTIMATES FURNISHED BY FORMULA (6)

An examination of Tables I and II points to the general fact that the estimates furnished by formula (6) tend to be in substantial error for high values in r_t (above ·60) and for larger values of k (above 1·0). When $r_t \leqslant \cdot 50$ and $k \leqslant \cdot 90$, the magnitude of error is less than 6 per cent. for all h except $h = \cdot 80$. For the range $\cdot 50 < r_t \leqslant \cdot 70$ the size of the error in general is under 10 per cent. for values of k less than ·80 or ·90 except when $h = 0$. In the instance that $h = 0$, however, errors associated with higher values in r_t are least for intermediate values of k (between ·40 and 1·1) and greatest when $k < \cdot 40$ and when $k > 1 \cdot 1$.

Similar statements may also be formulated concerning the percentage of error with respect to the size of the phi coefficient obtained. In general it is recommended that formula (6) should not be used for $h = 0$ and $h = \cdot 60$ when k is greater than unity and when $\varphi_{p'}$ exceeds ·50 or ·70 respectively. However, the entries in Table I representing percentages of error may be employed to advantage in that a corrected value of the coefficient may be obtained. In most instances use of linear interpolation should suffice to yield estimates of r_{tb} accurate to two digits. In fact the same degree of accuracy may be expected in the estimation of the values of certain desired parameters in Table I when magnitudes of other necessary parameters are known.

VIII. INTERRELATIONSHIPS AMONG PARAMETERS

Several interesting patterns of interrelationships among the parameters r_t, r_{tb}, $\varphi_{t'}$, k for either $h = 0$ or $h = \cdot 60$ are to be found upon examination of Table I. Perhaps two of the most evident features are (1) the decrease in the magnitudes of both r_{tb} and $\varphi_{p'}$ associated with an increase in the size of k (i.e., with an increase in the degree of departure of the level of item difficulty from the median value of $P = \cdot 50$) and (2) the approximate equivalence of r_t and $\varphi_{p'}$ when $h = \cdot 60$ for values of k representing moderate levels of item difficulty. These are but two among several patterns of interrelationships which are apparent within the tables.

IX. SUMMARY

A formula has been derived which furnishes an estimate of a point biserial coefficient of correlation from knowledge of the magnitude of a phi coefficient computed from use of contrasted groups of equal proportions. A technique is described by which the degree of systematic error in the formula may be ascertained from Pearson's *Volumes of Normal Bivariate Surface*. Corresponding to various magnitudes of the tetrachoric coefficient percentages of systematic error are presented in two tables for five different types of extreme groups (each of which consists of equal proportions of the total criterion group) as a function of selected levels of item difficulty.

In addition within the first table entries are available for the theoretical point biserial coefficient of the total criterion sample and, in the instance of two different sets of proportions of individuals in the contrasted groups, for the theoretical phi coefficient and the corresponding formula estimate of the point biserial coefficient. These two sets of proportions represent two familiar situations in item-analysis work in which the upper and lower halves and the upper and lower 27 percentages (approximately) of the total criterion sample are employed.

REFERENCES

1. Guilford, J. P. (1950). *Fundamental Statistics in Psychology and Education.* New York : McGraw-Hill.
2. Guilford, J. P. (1941). ' The phi coefficient and chi square as indices of item validity.' *Psychometrika*, VI, 11–19.
3. Guilford, J. P., and Perry, N. C. (1951). ' Estimation of other coefficients of correlation from the phi coefficient.' *Psychometrika*, XVI, 335–46.
4. Hertzka, A. F., Michael, W. B., and Perry, N. C. (1952). ' Systematic error in estimates of a phi coefficient.' *Educational and Psychological Measurement*, XII, 285–92.
5. Kelley, T. L. (1947). *The Kelley Statistical Tables.* Cambridge : Harvard University Press.
6. Michael, W. B., Perry, N. C., and Hertzka, A. F. (to appear). ' Abacs for rapid estimation of a tetrachoric coefficient of correlation from a phi coefficient calculated from use of extreme tails of a normal distribution of criterion scores.' *Educational and Psychological Measurement.*
7. Michael, W. B., Perry, N. C., and Hertzka, A. F. (to appear). ' The determination of the amount of systematic error in a formula for estimation of a coefficient of tetrachoric correlation from use of the extreme tails of a normal distribution of criterion scores.' *Educational and Psychological Measurement.*
8. Pearson, Karl (1931). *Tables for Statisticians and Biometricians*, Part II. England : Cambridge University Press.
9. Perry, N. C., and Michael, W. B. (1951). ' The estimation of a phi coefficient for an entire criterion group from a phi coefficient calculated from use of the extreme tails of a normal distribution of criterion scores.' *Educational and Psychological Measurement*, XI, 629-39.
10. Perry, N. C., and Michael, W. B. (in the press). ' The relationship of the tetrachoric coefficient to the phi coefficient estimated from use of the extreme tails of a normal distribution of criterion scores.' *Educational and Psychological Measurement.*

For Tables I and II see overleaf.

NOTE ON TABLE I.

Values corresponding to various coefficients of tetrachoric correlation r_t of the theoretically expected point biserial coefficient r_{pb} and phi coefficient $\varphi_{p'}$ and of $_{\varphi_{p'}}r_{pb}$ and percentages of systematic error 100ε arising from use of $_{\varphi_{p'}}r_{pb}$ as an estimate of r_{pb} for two sets of proportions of individuals p' ($p' = \cdot5000$ and $p' = \cdot2743$) in the upper and lower criterion groups as a function of item difficulty k.

TABLE I

Item Difficulty		$r_t = \cdot10$							$r_t = \cdot20$							$r_t = \cdot30$						
			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)				$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)				$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)		
k	P	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε
0·0	·5000	080	064	080	0·2	097	080	0·1	160	128	161	0·7	195−	160	0·4	239	194	243	1·6	294	242	1·0
0·1	·4602	080	064	080	0·2	097	080	0·1	159	128	160	0·7	194	160	0·4	239	194	243	1·5	293	241	1·0
0·2	·4207	079	063	079	0·2	096	079	0·1	158	127	159	0·7	193	159	0·4	238	192	241	1·5	291	240	0·9
0·3	·3821	078	063	079	0·1	095+	079	0·1	157	126	158	0·6	191	157	0·3	235+	191	239	1·4	288	237	0·7
0·4	·3446	077	062	078	0·1	094	078	0·0	155−	124	156	0·6	189	155+	0·2	232	188	236	1·3	284	234	0·5
0·5	·3085+	076	061	076	0·1	093	076	0·0	152	122	153	0·5	185+	153	0·1	229	185−	231	1·2	278	229	0·2
0·6	·2743	075−	060	075−	0·1	091	075−	0·0	149	120	150	0·4	181	149	−0·1	224	181	226	1·0	272	224	−0·1
0·7	·2420	073	058	073	0·1	089	073	−0·1	146	117	146	0·3	177	145+	−0·2	219	176	220	0·8	264	218	−0·5
0·8	·2119	071	057	071	0·1	086	071	−0·1	142	113	142	0·2	172	141	−0·4	213	171	214	0·5	256	211	−1·0
0·9	·1841	069	055−	069	0·0	083	069	−0·1	137	110	137	0·1	166	136	−0·7	206	165−	206	0·3	247	203	−1·5
1·0	·1587	066	053	066	0·0	080	066	−0·2	132	106	132	0·0	160	131	−0·9	199	158	199	0·0	236	195−	−2·0
1·1	·1357	064	051	064	0·0	077	063	−0·3	127	101	127	−0·2	153	126	−1·2	191	152	190	−0·4	226	186	−2·6
1·2	·1151	061	049	061	−0·1	074	061	−0·4	122	097	121	−0·3	146	120	−1·5	183	145−	181	−0·7	215−	177	−3·3
1·3	·0968	058	046	058	−0·1	070	058	−0·5	116	092	115+	−0·5	138	114	−1·8	174	137	172	−1·1	203	167	−3·9
1·5	·0668	052	041	052	−0·2	063	052	−0·6	104	082	103	−0·9	123	101	−2·5	156	122	153	−2·0	179	147	−5·4
1·7	·0446	046	036	045+	−0·3	055−	045+	−0·7	091	072	090	−1·3	107	088	−3·2	137	106	133	−3·0	155−	127	−6·9

Item Difficulty		$r_t = \cdot40$							$r_t = \cdot45$							$r_t = \cdot50$						
			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)				$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)				$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)		
k	P	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}r_{pb}$	% Error 100ε
0·0	·5000	319	262	328	2·9	395−	325−	1·8	359	297	372	3·7	446	367	2·3	399	333	418	4·7	498	410	2·8
0·1	·4602	319	261	328	2·8	394	324	1·7	358	296	372	3·7	445−	366	2·2	398	333	417	4·7	497	409	2·7
0·2	·4207	317	260	326	2·8	391	322	1·5	356	295−	369	3·6	442	363	2·0	396	330	414	4·5	493	406	2·4
0·3	·3821	314	257	322	2·6	386	318	1·2	353	291	365+	3·4	436	359	1·6	392	326	409	4·2	486	400	2·0
0·4	·3446	310	253	317	2·4	380	313	0·8	349	287	359	3·1	428	352	1·1	387	321	402	3·9	477	392	1·3
0·5	·3085+	305−	248	311	2·1	372	306	0·3	343	281	352	2·7	418	344	0·4	381	314	394	3·4	465+	383	0·5
0·6	·2743	299	243	304	1·8	362	298	−0·3	336	274	344	2·3	407	335−	−0·4	373	306	384	2·8	451	372	−0·5
0·7	·2420	292	236	296	1·4	351	289	−1·0	328	266	334	1·7	393	324	−1·3	365-	297	372	2·1	436	359	−1·6
0·8	·2119	284	228	286	0·9	338	278	−1·8	319	257	323	1·1	379	312	−2·3	354	287	359	1·4	418	344	−2·9
0·9	·1841	275−	220	276	0·4	325−	267	−2·7	309	248	310	0·5	363	298	−3·4	343	275+	345+	0·5	399	329	−4·2
1·0	·1587	265−	211	264	−0·2	310	255+	−3·6	298	237	297	−0·3	345+	284	−4·6	331	263	330	−0·4	379	312	−5·7
1·1	·1357	254	201	253	−0·8	295−	243	−4·7	286	226	283	−1·1	327	269	−5·9	318	250+	314	−1·4	359	295+	−7·2
1·2	·1151	243	191	240	−1·4	279	230	−5·7	274	214	269	−1·9	309	254	−7·2	304	237	297	−2·5	337	278	−8·8
1·3	·0968	232	181	227	−2·1	262	216	−6·8	261	202	253	−2·8	290	239	−8·6	290	223	279	−3·6	315+	260	−10·4
1·5	·0668	207	159	200	−3·7	229	188	−9·2	233	177	222	−4·8	251	207	−11·4	259	194	243	−6·2	272	224	−13·7
1·7	·0446	182	138	172	−5·5	196	161	−11·7	205+	152	191	−7·1	214	176	−14·3	228	166	208	−8·9	230	189	−17·1

TABLE I (*continued*)

Item Difficulty		$r_t = \cdot55$						$r_t = \cdot60$						$r_t = \cdot65$					
			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)		
k	P	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε
0·0	·5000	439	371	465−	5·9	552	454	3·4	479	410	513	7·2	606	499	4·1	519	450+	565−	8·9	661	544	4·9
0·1	·4602	438	370	464	5·8	550−	453	3·3	478	409	512	7·2	604	497	4·0	518	449	563	8·7	659	542	4·7
0·2	·4207	436	367	460	5·6	545+	449	3·0	475+	405+	508	6·9	598	492	3·6	515−	445+	558	8·4	652	536	4·2
0·3	·3821	432	363	454	5·3	537	442	2·4	471	400	501	6·5	588	484	2·8	510	439	550+	7·8	640	527	3·3
0·4	·3446	426	356	447	4·8	526	433	1·6	465−	393	492	5·8	575+	473	1·8	504	430	539	7·1	625+	514	2·1
0·5	·3085+	419	348	437	4·2	512	421	0·5	457	383	481	5·1	559	460	0·6	495+	419	526	6·1	606	499	0·7
0·6	·2743	411	339	425−	3·4	496	408	−0·7	448	372	467	4·1	540	444	- 0·9	485+	406	509	4·9	584	480	−1·1
0·7	·2420	401	328	411	2·6	477	393	−2·0	437	360	451	3·1	518	427	−2·5	474	392	491	3·6	559	460	−3·0
0·8	·2119	390	316	396	1·6	457	376	−3·5	425+	346	433	1·8	495−	407	−4·3	461	375+	470	2·0	531	437	−5·1
0·9	·1841	378	303	380	0·5	435+	358	−5·2	412	330	414	0·5	470	386	−6·2	446	357	448	0·4	503	414	−7·3
1·0	·1587	364	289	362	−0·6	412	339	−6·9	397	314	394	−1·0	443	365−	−8·2	430	339	424	−1·4	473	389	−9·6
1·1	·1357	350	274	343	−1·9	388	319	−8·7	382	297	372	−2·5	416	343	- 10·3	414	319	400	−3·4	442	364	−12·0
1·2	·1151	335−	258	324	−3·3	364	299	−10·5	365+	279	350−	−4·2	389	320	−12·4	396	299	374	−5·4	412	339	−14·3
1·3	·0968	319	242	304	−4·7	339	279	−12·4	348	261	327	−5·9	361	297	−14·5	377	278	349	−7·4	381	314	−16·7
1·5	·0668	285+	210	263	−7·7	290	239	−16·2	311	225−	281	−9·6	307	253	−18·8	337	237	298	−11·8	322	265−	−21·4
1·7	·0446	251	178	223	−11·0	244	201	−20·0	273	189	237	−13·5	256	211	−22·9	296	198	248	−16·2	267	219	−25·9

Item Difficulty		$r_t = \cdot70$						$r_t = \cdot75$						$r_t = \cdot80$					
			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)			$h = 0{\cdot}0$ ($p' = \cdot5000$)			$h = 0{\cdot}6$ ($p' = \cdot2743$)		
k	P	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	r_{pb}	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε	$\varphi_{p'}$	$\varphi_{p'}{}^{r_{pb}}$	% Error 100ε
0·0	·5000	559	494	619	10·8	718	591	5·7	598	540	677	13·1	775+	638	6·6	638	590	740	15·9	834	686	7·5
0·1	·4602	558	492	617	10·6	715−	588	5·5	597	538	674	12·9	772	635+	6·4	637	588	737	15·7	830	683	7·2
0·2	·4207	554	487	611	10·2	707	582	4·9	594	532	667	12·3	762	627	5·6	634	581	728	14·9	819	674	6·3
0·3	·3821	549	480	601	9·5	693	571	3·8	589	523	656	11·4	747	615−	4·4	628	570	714	13·7	801	659	4·9
0·4	·3446	542	470	588	8·5	675+	556	2·4	581	511	640	10·1	725+	597	2·7	620	554	694	12·0	776	639	3·0
0·5	·3085+	534	457	572	7·2	653	537	0·7	572	495−	620	8·5	699	576	0·7	610	535−	670	9·9	746	614	0·7
0·6	·2743	523	441	553	5·7	627	516	−1·3	560	477	597	6·6	670	551	−1·6	598	512	642	7·5	712	586	−1·9
0·7	·2420	510	424	531	4·0	598	492	−3·6	547	456	571	4·5	637	524	−4·2	583	488	611	4·8	675−	555+	−4·8
0·8	·2119	496	404	507	2·1	567	467	−6·0	532	433	543	2·1	601	495−	−6·9	567	461	578	1·8	635−	522	−7·9
0·9	·1841	481	384	481	0·8	534	440	−8·5	515−	409	513	−0·4	565−	465−	−9·8	549	433	542	−1·3	594	489	−11·0
1·0	·1587	464	362	454	−2·1	501	412	−11·1	497	384	481	−3·1	527	434	−12·7	530	404	506	−4·5	552	454	−14·3
1·1	·1357	445+	340	426	−4·4	467	384	−13·8	477	358	449	−5·9	489	403	−15·6	509	375−	469	−7·8	511	420	−17·4
1·2	·1151	426	317	397	−6·8	433	356	−16·4	456	333	417	−8·7	452	372	−18·5	487	345+	433	−11·1	470	387	−20·6
1·3	·0968	406	294	368	−9·3	399	329	−19·0	435−	307	385−	−11·5	416	342	−21·3	464	317	397	−14·3	430	354	−23 6
1·5	·0668	363	248	311	−14·3	335+	276	−24·1	389	257	322	−17·2	346	285+	−26·7	415−	263	330	−20·6	356	293	−29·4
1·7	·0446	319	205+	257	−19·3	276	227	−28·9	342	211	264	−22·7	283	233	−31·8	365−	214	268	−26·4	289	238	−34·7

TABLE II

Percentages of systematic error 100ε corresponding to various coefficients of tetrachoric correlation r_t at selected levels of item difficulty k arising from use of $\varphi_{p'}r_{pb}$ as an estimate of r_{tb} for three sets of proportions of individuals in the extreme groups of a total criterion sample : $p' = \cdot3085$ ($h = \cdot50$), $p' = \cdot2420$ ($h = \cdot70$), and $p' = \cdot2119$ ($h = \cdot80$).

Coefficients of Tetrachoric Correlation		$r_t = \cdot10$			$r_t = \cdot20$			$r_t = \cdot30$			$r_t = \cdot40$			$r_t = \cdot45$			$r_t = \cdot50$		
Item Difficulty		$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$
k	P																		
0·0	·5000	0·1	0·1	0·1	0·5	0·3	0·2	1·2	0·8	0·5	2·1	1·4	0·9	2·7	1·7	1·1	3·4	2·1	1·4
0·3	·3821	0·1	0·1	0·0	0·4	0·2	0·1	0·9	0·5	0·2	1·6	0·8	0·3	2·1	1·0	0·4	2·6	1·2	0·4
0·6	·2743	0·0	−0·1	−0·1	0·1	−0·2	−0·3	0·1	−0·5	−0·8	0·2	−0·8	−1·4	0·2	−1·1	−1·8	0·3	−1·4	−2·3
0·9	·1841	−0·1	−0·2	−0·3	−0·5	−0·8	−1·0	−1·1	−1·9	−2·3	−2·0	−3·4	−4·1	−2·6	−4·3	−5·1	−3·3	−5·2	−6·2
1·3	·0968	−0·4	−0·5	−0·6	−1·5	−2·1	−2·3	−3·4	−4·5	−5·1	−6·0	−7·7	−8·6	−7·6	−9·6	−10·6	−9·3	−11·5	−12·6
1·7	·0446	−0·7	−0·9	−1·0	−2·8	−3·6	−4·0	−6·2	−7·6	−8·3	−10·7	−12·7	−13·6	−10·8	−15·4	−16·4	−15·9	−18·2	−19·3

Coefficients of Tetrachoric Correlation		$r_t = \cdot55$			$r_t = \cdot60$			$r_t = \cdot65$			$r_t = \cdot70$			$r_t = \cdot75$			$r_t = \cdot80$		
Item Difficulty		$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$	$h = \cdot5$	$h = \cdot7$	$h = \cdot8$
k	P																		
0·0	·5000	4·2	2·6	1·6	5·1	3·1	1·9	6·1	3·6	2·0	7·2	4·0	2·2	8·5	4·5	2·1	9·9	4·8	1·8
0·3	·3821	3·2	1·5	0·5	3·8	1·7	0·5	4·5	2·0	0·5	5·3	2·2	0·4	6·2	2·4	0·7	7·2	2·5	−0·1
0·6	·2743	0·3	−1·7	−2·8	0·3	−2·0	−3·3	0·2	−2·4	−3·9	0·1	−2·9	−4·5	0·0	−3·3	−5·1	−0·1	−3·8	−5·6
0·9	·1841	−4·1	−6·3	−7·4	−5·0	−7·4	−8·7	−6·0	−8·6	−10·0	−7·1	−9·9	−11·2	−8·4	−11·1	−12·4	−9·7	−12·2	−13·3
1·3	·0968	−11·2	−13·6	−14·7	−13·2	−15·7	−16·8	−15·5	−17·9	−18·9	−17·8	−20·0	−20·9	−20·3	−22·1	−22·8	−22·9	−24·2	−24·5
1·7	·0446	−18·8	−21·1	−22·1	−21·7	−24·0	−24·8	−24·8	−26·8	−27·5	−28·0	−29·5	−30·0	−31·1	−32·2	−32·4	−34·2	−34·8	−34·7

Vol. V
Part III

The British Journal of Psychology,
Statistical Section

November,
1952

THE HIERARCHICAL ORGANIZATION OF COGNITIVE LEVELS

By E. M. MOURSY
Institute of Education, Cairo

I. *History of Problem.* II. *Alternative Hypotheses and Methodological Assumptions.* III. *Plan of Experiment :* (1) *Tests,* (2) *Subjects.* IV. *Analysis of Results.* V. *General Inferences.* VI. *Summary and Conclusions.*

I. HISTORY OF PROBLEM

Problem. The general problem with which the following investigation is concerned is the classification of mental abilities. Its more immediate purpose is to verify or refute, by means of a statistical analysis, the hypothesis that cognitive processes can be classified into a hierarchical scheme, that is, into a ramifying system of distinct developmental levels, each successive level being characterized by an increasing complexity and by a special quality of its own. As there is still some misunderstanding about the precise nature of such schemes, I shall preface the account of my own experiments by an attempt to examine the exact meaning of the phrase, as used in statistical psychology.[1]

" The hierarchical theory," says Vernon in a recent volume, " was first put forward by Burt under the influence of McDougall " ((16), p. 25). In vaguer forms the notion may be found in many earlier speculations ; and, during the last few years, it has been accepted by an increasing number of experimental and statistical writers. Unfortunately the way such writers use the term appears to imply quite different interpretations of the underlying idea, some of which may even be mutually conflicting. On the other hand, many psychologists have, implicitly or explicitly, altogether rejected the notion. Accordingly, there seems an urgent need to re-examine the whole subject.

History. In order to formulate in more precise terms the main issues involved, it will be helpful to begin with a brief historical review of the theories actually put forward by various psychologists who have discussed the problem in the past. As Burt points out, the doctrine that the mind may be regarded as possessing a hierarchical structure has been reached by four different lines of approach : (1) the introspective analysis of mental processes ; (2) the comparative study of their evolu-

[1] I am the more encouraged to do so because the editors have recently expressed a wish for articles of an expository type. The specialist will no doubt prefer to go straight to the statistical sections reporting the results of the present research (Sect. III *et seq.*).

On the general nature of ' hierarchical ' and other types of classification, see Sigwart, C., *Logic,* 1895, Vol. II, ' Methodology,' esp. pp. 9, 158f., 508f., and Stebbing, S., *A Modern Introduction to Logic,* 1930, pp. 439f. In this article the term ' hierarchy ' will be used in accordance with the definition proposed by Burt (this *Journal,* II, p. 63). In *Analysis Situs* such relational schemes are more commonly called ' trees.' A similar but slightly broader definition is given by Woodger in the notation of *Principia Mathematica* (Woodger, J. H., *The Axiomatic Method in Biology,* 1937, p. 42). There are two limiting cases. (i) If all branches except the central trunk are omitted, the ' hierarchical order ' is reduced to a vertical or ' serial order ' : (in Spearman's writings the term usually has this restricted meaning). (ii) If all rows except the last are omitted, the scheme is reduced to a set of co-ordinate species with unspecified relations.

tionary development ; (3) the investigation of their neurological bases ; and (4) the statistical comparison of their similarities and differences.[1]

(1) *Introspective Classification.* In its broader outlines the conception appears to have arisen from early attempts to introduce some kind of logical order into the chaotic variety of observable mental activities, and of the numerous ' faculties ' and ' propensities ' postulated by philosophical writers in their attempts to describe the more obvious phenomena of ordinary everyday life. The simplest schemes are at least as old as early Hellenic speculations on " the many and the one." [2] " In the more elaborate scheme of Plato," says Sully, " the mental powers are arranged in a hierarchy. . . . A similar view of the mind as a graded series of powers reappears in Aristotle, only here it becomes incorporated into a genetic view of mind, and so prepares the way for modern developmental theory." [3]

Starting with the distinctions introduced by Aristotle, a schematized classification into two main aspects (intellectual and appetitive) and four main levels (special senses, the common sense, memory, and reasoning) was gradually worked out in the writings of scholastic philosophers, notably St. Thomas Aquinas (cf. (13)). It was based chiefly on introspective analyses ; and culminated in the somewhat complicated tabulations of the faculty school. The best known version of the faculty doctrine is that printed in the ' synoptic table ' given by Wundt, and taken (as he says) from " the last and most influential of the systematizers," Christian Wolff.[4] This shows two main aspects (distinguishing ' cognitive capacities,' *Erkenntnisvermögen*, from ' capacities of desire,' *Begehrungsvermögen*) and two main levels (a ' lower ' or perceptual level and a ' higher ' or conceptual).

The early nineteenth century writers changed the scheme from a cross classification to a branching classification, of the kind to which the logicians of that date applied the term ' hierarchy ' in a more technical sense.[5] Borrowing Linnæus' taxonomic procedure, they sought to divide and subdivide mental abilities into genera, species, subspecies, and so forth.[6] Probably the most familiar of these modified versions is that popularized by phrenological writers like Spurzheim and Combe, who started with a broad division of ' mental powers ' into two main ' orders ' (intellectual faculties and emotional propensities), and then subdivided each into ' families,' ' genera,' ' species,' and ' varieties.' [7]

(2) *Evolutionary Classification.* When the Linnæan classification, in the more general fields of botany and zoology, was interpreted by evolutionary writers like Darwin and Spencer as the outcome of the progressive development of the plant and animal kingdoms, it was natural to account for the differentiation of mental activities by a similar process. According to Spencer the ' levels ' of older writers are to be regarded as ' developmental stages ' ; and it is he who is responsible for explicitly introducing the notion of a ' concrete hierarchical organization ' into the description of the mental activities of the individual, and explaining them by reference to a similar organization of ' afferent ' and ' efferent nervous arcs.' The ' arcs ' are assumed to be combined into ' centres ' ; and the centres are said to be anatomically arranged in levels or stages of increasing ' superordination,' the ' lower ' being subordinated to the ' higher.'

Spencer is the first psychological writer who actually uses the term ' hierarchy ' in this connexion. His own classification accepts the broad distinction between ' cognitions ' and ' feelings,' and then divides each aspect into two main classes—' presentative ' and ' representative.' Each of these in turn is subdivided into two subclasses, so that there are four main levels in all. The scheme so reached on the cognitive side, with the processes specified by Spencer as typical, is reproduced below :—

I. *Presentative Cognitions :*
  i. Presentative simply, e.g., sensation ;
  ii. Presentative-representative, e.g., perception.

[1] Here and in much of what follows I am indebted, not only to the general histories of Brett and Murphy, but also to the accounts of the early development of factorial methods given by William Brown (*Mental Measurement*, 1912), Thorndike (*Educational Psychology*, 1914), P. E. Vernon (*The Structure of Human Abilities*, 1950), and Cyril Burt in an unpublished *Historical Note* which he has been good enough to send me.

[2] See Siebeck, H., *Geschichte der Psychologie*, 1880, Vol. I.

[3] Sully, J., *The Human Mind* (1898), II, Appendix A, ' The Division or Classification of Mental Phenomena,' p. 327.

[4] *Grundzüge d. Physiol. Psychologie*, 1874 (6th ed., 1908), I, p. 17.

[5] For the biological conception of ' hierarchies of levels of organization ' and ' the implications of this notion for biological explanation,' see more particularly Woodger, J. H., *Biological Principles*, 1929, ch. VI, sect. 12.

[6] See the various examples cited by Bain in his Appendix on ' Classifications of the Intellectual Powers ' (*Sense and Intellect*, 2nd ed., 1864, pp. 539–640).

[7] Quoted by Burt in *The Measurement of Mental Capacities* (Oliver and Boyd, 1926, pp. 9–10), who compares ' modern factors with old-fashioned faculties.' See also Allport, G. W., *Personality* (1938, pp. 84f.), who also draws a similar analogy (p. 81).

II. *Representative Cognitions :*

iii. Representative simply, e.g., recollection and represented sensations ;
iv. Re-representative, e.g., abstract thought.

Although the terms chosen suggest qualitative distinctions, the gradation from one class to another, he insists, is perfectly continuous ; it is due primarily to increasing differentiation and integration which results in an increasing complexity of the total process, the whole developing ' like the branches of a tree.' [1]

The unitary ability which differentiates in this way he terms ' intelligence.' Intelligence, we are told, consists essentially in " a power of combining many impressions." Throughout the evolution of the animal kingdom and the development of the individual organism, mental activity, as an effect of this progressive integration of progressively differentiated tendencies, exhibits " a decidedly serial arrangement in proportion as intelligence advances." Each active adjustment made by the individual is " the *resultant* of many tendencies acting together ; the next state is produced by composition of forces " (his italics). And the aim of the psychologist's ' analysis ' is " to *resolve* each concrete type of cognition into its *components*." [2]

(3) *Neurological Evidence.* Spencer's somewhat speculative account was taken over and developed by the leading neurologist of the late nineteenth century, Hughlings Jackson,[3] and supported by a wealth of anatomical and clinical evidence in his description of the structure and functions of the central nervous system. His conclusions are concisely summarized by Schäfer, as follows : " The term ' level ' was adopted by Hughlings Jackson to imply stages of complexity in the sensori-motor functions : simple reflex actions (in the spinal cord or bulb) form the *first* or lowest level ; above these come centres in the sensory or motor areas of the cortex, the action of which is somewhat more complex and attended by consciousness : these form the *second* level ; the *third* or highest level is formed by those centres or paths in the associative areas of the cortex which are conceived to be the seat of memory and of ideation." [4] It will be noted that Spencer's two highest levels (for memory and for ideation) are here merged into one, and a reflex level is added to provide a foundation for the upper storeys. The whole thus forms " a hierarchy of three superimposed sensori-motor levels " ; and the term ' hierarchy ' now includes literal government or control (ἀρχή). Sherrington expresses himself in terms of a similar conception, though his account is far less schematic (2) ; and McDougall has developed the Jacksonian scheme in still greater detail to supply " a working model of the structure of the mind." [5]

(4) *Statistical Evidence.* A new approach to such problems was opened up in the investigation undertaken by Clark Wissler at the suggestion of J. M. Cattell. The novelty consisted in applying the correlation technique developed by Galton and Pearson to quantitative data collected by Galton's device of mental tests ; and it was " hoped thereby to find a means by which the fundamental elements of *general* and *specific abilities* could be isolated and valued." [6] Later followers of Galton and Pearson, such as Thorndike, Burt, and Spearman (the last two directly influenced by McDougall's suggestions), made considerable use of the Spencerian doctrine of a ' hierarchy of intelligences ' in

[1] *Principles of Psychology*, 2nd ed., 1870, I, Pt. i, ' The Structure of the Nervous System ' (esp. p. 30), Pt. iii, ch. 7, ' The Correspondence as Increasing in Complexity ' (pp. 350f.), Pt. iv, ch. 3, ' The Growth of Intelligence ' (pp. 418f.), and II, Pt. viii, ch. 2, ' Classification ' (pp. 513–17).

[2] *Ibid.*, I, Pt. i, ch. 1, ' The Nature of Intelligence,' and ch. 2, ' The Law of Intelligence.' The terms selected to describe the several levels appear to have been suggested by Wolff's use of *vis repraesentativa* in a similar way. ' Degree of Representativeness,' it is said, forms " a measure of *complexity*, and implies a proportionate degree of *integration*," and so serves also " as a measure of evolution or development " (II, pp. 516–17 : Spencer's italics).

[3] ' On the Relations of the Divisions of the Nervous System,' *Brit. Med. Journ.*, 1898, I, pp. 65f. ; also *Selected Writings*, II, pp. 63f. A more detailed account will be found in (13). The psychological student of to-day will be familiar with the general nature of the scheme from such diagrams as that given in Boring, Langfeld, and Weld, *Foundations of Psychology*, 1948, p. 34, Fig. 16.

[4] Schäfer, E. A., *Textbook of Physiology*, II, pp. 697–8 : (quoted by Burt).

[5] McDougall, W., *Psychology : The Science of Behaviour*, 1912, ch. III, ' The Structure of the Mind ' ; cf. also (1), 1905, ch. I, where the evidence for this application of Jackson's view is briefly summarized. McDougall adds that " it is probable that future research may enable us to distinguish a hierarchy of (subordinate) levels among those grouped to constitute the third (or highest) level " (p. 23). In later writings he develops this last suggestion, claiming that, in the course of mental evolution and development, " new abilities are not only formed by gradual differentiation out of pre-existing abilities, but also, when differentiated, continue to function as part of larger systems of which the parent ability remains an essential feature." And he now points out (1932) that, although originally suggested by " the comparative study of animal and human activities," this picture of the mind's structure proved to be " in agreement with, and supported by, the analysis of mental functions achieved by the testing and statistical examination of the data " ((11), pp. 93, 95, 99).

[6] Wissler, C., ' The Correlation of Mental and Physical Tests,' *Psych. Mon. Sup.*, III, 1901, pp. 1–62.

their various correlational studies. But they were led to very different interpretations of the fundamental scheme.

Three main versions may be distinguished ; and it will be part of our task to obtain, if possible, empirical evidence deciding in favour of one or other of the three.

(i) *Both General and Group Factors.* What may be called the original version, that outlined by Spencer and amplified by neurologists like Jackson and physiological psychologists like McDougall, is represented by the early researches of Burt and his collaborators (3), (4), (6).

Burt and Moore suggest " four criteria by which we may judge whether one process belongs to a higher or lower level : first, their supposed order of evolution ; secondly, their supposed importance for intellectual and practical life ; thirdly, their apparent complexity ; fourthly, their apparent correlation with the general factor of intelligence." This leads them to " arrange processes to be tested in four main classes : (*a*) simple perceptual and motor, (*b*) complex perceptual and motor, (*c*) associative, and (*d*) relational or reasoning." [1] And their investigations were planned so that " the tests should represent *all* the main aspects and levels of mental process " : previously, " investigations on the relations between mental abilities (Wissler's and Thorndike's for example) had been characterized by one serious defect : the tests were not selected on any systematic plan " ((4), p. 94).

Their results appeared to confirm the conception of an integrative structure in the mind, with a series of abilities or ' factors ' of varying breadth, divided and subdivided into four main levels. " The processes of the lowest level consist of simple sensory or motor processes, such as can be isolated and measured by tests of sensory ' thresholds ' and the timing of ' simple reactions ' ; the next level includes the more complex processes of perception and co-ordinated movement, as in experiments on the apprehension of form and pattern or on ' compound reactions ' ; the third is the associative level—the level of memory, imagery, and habit formation ; the fourth and highest level of all involves the apprehension and application of relations. Intelligence, as the integrative capacity of the mind, is manifested at every level, but these manifestations differ not only in degree, but also in their qualitative nature." [2] The analysis into ' factors ' is conceived as a resolution of concrete manifestations into components ; but equally the ' factors ' can, for purposes of computation, themselves be treated as ' resultants.' Hence the familiar formulæ for the resolution and composition of forces in dynamics become available (at least as first approximations) for the analysis and reconstruction of mental tendencies in the study of the individual mind.

The distinctive feature of this version is that it involves a recognition of *both* general *and* specific abilities. In describing his original problem Burt ((3), p. 95) refers to Meumann's recent work on mental tests, which (without any statistical confirmation) had drawn a distinction between (*a*) " sogenannte *allgemeine Fähigkeiten,* die sich bei jeder geistigen Arbeit mit betätigen," and (*b*) " *spezieller Fähigkeiten,* oder Partialfähigkeiten (in dem Sinne der üblichen Klassificationen der Gegenwartigen Psychologie), die sich nur an bestimmten Arten von Bewusstsein betätigen, d.h., Sinneswahrnehmung, Gedächtnis, beziehende Denken, usw." [3] On the other hand, the associationists (as well as several of their opponents—Ward, for example) still maintained that any such doctrine was reminiscent of the discredited theory of faculties, and insisted on only a *single* fundamental cognitive process. In order to verify the presence of these more specialized factors peculiar to certain levels or aspects (so Burt maintained) a suitable statistical technique was required. Accordingly, the *theoretical* correlations expected as a result of the general factor were systematically computed and subtracted from the *observed* correlations obtained by actual experiment. The significance of the discrepancies was then tested by the appropriate probable errors. Since in the earliest investigations all the tests were administered individually, only a small number of children could be tested in each experiment. Nevertheless, it was found that there was " a small but discernible tendency for subordinate groups (of processes) to correlate together," and in the later investigations, based on tests applied

[1] Burt, C., and Moore, R. C., *J. Exp. Ped., loc. cit. sup.*, I, p. 363. To these four criteria Burt later added two others (which in historical order should perhaps be placed second and third), namely, (*e*) the way in which a kind of devolution, proceeding by levels or stages, appears in cases of neurological disease and organic psychosis, and in the effects of drugs, fatigue, natural and hypnotic sleep, and the like, and (*f*) the way in which the histological study of the central nervous system has indicated its organization in levels or relays of superposed nervous arcs (no doubt far more complicated than the schematic four-level theory of the elementary text-books would suggest). It was more particularly facts under these two headings that influenced neurologists like Jackson, Sherrington, and Head to adopt the Spencerian theory : (cf. Jackson, H., ' Evolution and Dissolution of the Nervous System,' *Lancet*, I, 1884, pp. 555–8, 649–52, 739–44, quoted by Carmichael, L., *ap.* Stevens, S. S., *Handbook of Experimental Psychology*, 1951, pp. 288–301).

[2] Cf. (15), p. 104. The levels and tests employed by Burt and Flugel are set out in tabular form in (3), 1909, pp. 98–9 ; but tests of the highest level were not actually used till the later experiments (4), (5). The fuller set of tests used in the work with Moore is tabulated in the section on ' Experiments : Intellectual and Practical Activities ' in *J. Exp. Ped.*, I, 1912, p. 364.

[3] Meumann, E., *Experimentelle Pädagogik*, 1907, I, pp. 74f.

to much larger numbers, it was claimed that the presence of *group factors*, in addition to the ' highest common factor,' was fully established.

(ii) *General Factor Only.* Spearman, influenced, it would seem, partly by Ward's doctrine of a single cognitive ability, and still more by Spencer's reiterated statements that all the cognitive processes, from the simplest to the most complex, are differentiations of a single capacity called 'intelligence,' questioned from the outset the very existence of anything like ' special abilities ' limited to tests belonging to a special type or group. He maintained that the traditional conception of the mind could be greatly simplified by discarding the so-called ' special abilities ' or ' faculties.' For him, therefore, " the hierarchy of intelligences," as he termed it, consisted merely in a *linear* series formed by an increasingly efficient capacity for ' general discrimination.' [1] All the various cognitive abilities were, in his view, merely manifestations of one underlying function—' general intelligence ' ; and hence might be expected (in Spencer's phrase) to exhibit a uniform ' serial arrangement in proportion as intelligence advances.' [2] Thus, as he later contended in a joint paper with Hart (7), the organization of the mind is not ' multifocal ' but ' unifocal.' To support his contention, he reprints Burt's main table, and analyses it by a fresh procedure in the hope of demonstrating that, here as elsewhere, the result " is exactly that required by the *unifocal* theory of a General Common Factor." [3]

To show what would be the effect of superposing group factors, representing different mental levels, on a basic set of correlations due to a general factor only, Spearman and Hart present a theoretical correlation table in symbolic form (p. 57), and contend that the clusters of disproportionately high correlations that we should expect if additional group factors were present are *not*, in point of fact, discernible in the tables obtained from actual experiments.[4] This, of course, touches the real issue which we shall have to examine in the light of our own observations.

It will be seen that Spearman's attempted simplification was tantamount to a reversion to the older doctrine of serial development based on the idea of a unilinear *scala naturæ* in place of the more recent doctrine of divergent development, based on a branching *arbor naturæ*, which the term

[1] The elevation of ' discrimination ' to the role of the fundamental capacity (rather than ' apperception ' or ' attention,' as maintained by Wundt and Ward)—i.e., of analysis or differentiation rather than synthesis or integration—strikes the modern reader with some surprise. But it should be remembered that this suggestion had already been put forward by Bain, whose writings continued to represent the orthodox view in Great Britain until they were displaced by the alternative views of Ward and Stout: at the very outset of his text-book Bain declares that Discrimination is the ' fundamental property ' of ' Intelligence or Cognition ' : (cf. *Senses and Intellect*, pp. 5–7). Sully (head of the department in which Spearman obtained his first post) similarly held ' discrimination ' to be fundamental. This view was explicitly accepted by Spearman in his 1904 paper ; and in his joint paper with Hart he emphatically maintains that those authors who, like Burt, identify the general factor with ' synthetic power ' or capacity for integration are following a ' false track ' ((7), p. 71, and footnote).

[2] Spencer's alternate emphasis on integration and differentiation leads to statements that at first sight seem somewhat inconsistent. But there can be little doubt that he did not intend his readers to understand that the development of mental processes consisted *merely* in a single ' serial arrangement.' Indeed, he begins his chapter on classification of mental phenomena by comparing " the great groups, the subgroups, the sub-subgroups, etc.," to " a tree, *each* branch of which bears secondary and tertiary branches, with a continual divergence and re-divergence of the branches " (*op. cit.*, II, p. 510). This warning seems to have been deliberately prefixed to the classification as originally conceived (*Essays*, I, 1868, p. 321), in order to avoid the serial interpretation ; and the classification itself is said to be merely " roughly indicated and capable of expansion." Accordingly, in attributing a *purely* ' monarchic ' view to Spencer (*Abilities of Man*, pp. 4–6) Spearman scarcely does full justice to Spencer's conception : that was not ' monarchical ' but ' hierarchical.'

[3] (7), pp. 54f. In place of the method adopted by Burt (examining the differences between the observed correlations and the expected values required by the unifocal theory) he substitutes a criterion based on the intercolumnar correlations. The new procedure did not escape criticism ; and Spearman later proposed his tetrad difference criterion, which included a formal test of significance based on the probable error.

[4] In Burt's 1909 table, reprinted by Spearman, the ' clusters of disproportionately high correlations ' (as Burt himself admitted) did not, when tested by the probable error, show very conclusive evidence of group factors, because the children tested in his first experiments had been so few. But by the time Spearman and Hart published their paper, Burt (in conjunction with R. C. Moore and others) had already produced far more convincing evidence, based on data obtained with group tests. William Brown's correlations pointed in the same direction, when the significance of various differences was tested in the recognized way. However, Spearman's contention was that the usual significance tests were not appropriate in the present case and that his new intercolumnar criterion alone could yield a satisfactory pronouncement on the essential issue.

' hierarchy ' was intended to designate. Indeed, in later publications, he proposed the term ' monarchic ' to describe the ' unifocal ' theory, in contrast with the ' hierarchic ' or ' oligarchic.' [1]

(iii) *Group Factors Only.* While Spearman wished to dispense with the *special* abilities (represented by ' group factors '), Thorndike, on the other hand, preferred to omit the *general* ability (represented by the ' general factor '). In place of the traditional scheme, he proposed to substitute a division of cognitive abilities into three main levels only, (i) a level of Sensory Discrimination, (ii) a level of Association, and (iii) a level of Dissociation, i.e., abstraction and generalization ((5), pp. 191f.). This view and its implications were criticized and rejected in the joint paper by Hart and Spearman ((7), pp. 52f.). In a later publication, Thorndike, who also reprints Burt's table, sets forth his own reply, and endeavours to refute the interpretation given by Spearman. " That doctrine," he says ((8), p. 364), " requires not only that all branches of intellectual activity be positively correlated, which is substantially true, but also that they be bound in *all* cases by only *one* common factor, which is false. . . . Some traits (those of the same level) are closely related to each other, but their common element is only very loosely related to that of other groups." Nevertheless, Thorndike now appears less inclined to admit the influence of factors representing the *form* of mental processes (" power to discriminate, associate, reason, . . . " etc.), and to lay more stress on factors of *content* (verbal, numerical, spatial, and the like).[2]

Later still, in *The Measurement of Intelligence*, he appears to abandon any attempt to classify mental processes into sharply demarcated levels or classes. He still speaks of differences in ' level.' But he now uses the term to mean the relative difficulty of a composite test ; and the differences in difficulty are conceived as forming a continuous, not a discrete, series. " The orthodox view of intellect," he says, " has been that it is divided rather sharply into a lower half and a higher half. . . . The hypothesis which we now present asserts that the higher forms depend on the *same sort* of connexions as the lower, but require *many more* . . . in short, that *quality* of intellect depends merely on *quantity* of connexions." [3] He agrees, however, that there is a definite ' specialization within levels,' pointing out, for instance, that " there are individuals who are much better in arithmetical tasks than in linguistic tasks." And he thus still recognizes the influence of what other writers would call ' group factors for content.' [4]

Kelley, Thurstone, and their followers appear to keep rather to Thorndike's earlier view. They are sceptical of anything like a general factor, but recognize a number of group factors, which they identify with irreducible ' primary ' or ' unitary ' abilities. These, however, are apparently conceived as a miscellaneous collection of semi-independent faculties, not as forming an intelligible system developing or differentiating in a recognizable order. Some such view as this would at the moment appear to be the orthodox doctrine in America. Thus in the latest edition of Boring, Langfeld, and Weld's well-known text-book, we are told that " the principal result of work with factor analysis is that the tested abilities form groups or clusters because they involve common group factors ; the most clearly demonstrated are the verbal, numerical, and spatial "—i.e., content factors.[5]

[1] Cf., for instance, *Abilities of Man*, ch. I, ' The Monarchic Doctrine : Intelligence.' Chapter VI (describing an ' eclectic doctrine ') suggested at the time that Spearman was moving towards a doctrine of three factors—a general factor, a few broad group factors, and a number of specific or unique. But his subsequent explanations indicated that this interpretation was incorrect : (cf. also *ibid.*, p. 241, where we are told that " group factors are astonishingly rare " and usually attributable to " past experience rather than to native aptitude "). In his later writings we hear no more of " the hierarchy of intelligences " in the concrete sense, and the term ' hierarchy ' (placed in inverted commas) is used merely to designate the " serial arrangement " of the rows and columns of the correlation matrix, resulting from the serial development of various manifestations of intelligence as such.

[2] The change is doubtless due in part to the results of the experiments on ' formal training ' with which Thorndike had been largely engaged during the interval. In a footnote (p. 364) he also points out that, in a section added apparently as an afterthought, Spearman and Hart (not quite consistently) leave themselves a loophole : although in experimental work the so-called specific abilities " exercise a surprisingly small influence," nevertheless (they say) this may not be so " in ordinary life." They even admit that " the specific correlation discovered by Burt " (in the same table as they had already examined) lends some colour to his notion of a group factor for associative processes—a factor whose " range might cover all the diverse activities coming under the head of memorization." If so, the ' two-factor theory ' (general and specifics only) breaks down, since the group factor for memory would provide a striking instance of what Burt had called the third kind of factor.

[3] (9), ch. VI, ' Levels of Intellect ' : cf. also pp. 415f. By itself this statement would seem almost to imply a reversion to the simpler theory attributed to Spencer.

[4] Cf. *loc. cit.*, Index, *s.v.*, ' specialization ' : also p. 390.

[5] Boring, Langfeld, and Weld, *Foundations of Psychology*, p. 412. Here the traditional names of abilities appear as chapter headings—learning, recollecting, imagining and thinking, perception, sensation, etc. ; but the order is neither logical nor genetic. On the other hand, neurological evidence continues to confirm the doctrine of ' levels.' Thus (to mention only work on the cortex) de Crinis, examining

Nevertheless, on both sides of the Atlantic there has been of late a renewed and an increasing interest in the conception of a hierarchical organization of mental capacities, particularly among writers interested in individual psychology. In America the conception has been adopted, a little tentatively perhaps, by Hollingworth : in contrast to Thorndike he maintains that intelligence " involves *both* ' number of connexions ' *and* integration," and he describes mental activity as dependent on a ' hierarchy of adjustments ' arising from a ' hierarchy of potencies.' [1] Tolman speaks of ' hierarchies of demands,' and Clark Hull's ' third postulate ' describes organisms at birth as possessing the potentiality for a ' hierarchy of responses ' on which can be built up a ' hierarchy of habit-families.' [2] Allport has put forward " the conception of a hierarchy of levels produced by integration," which he believes may provide " much help in understanding the development of personality." [3] More recently still Piaget has described the mental development of the child as characterized by the appearance of a ' hierarchy of operations,' proceeding by developmental stages or ' levels ' very similar to those outlined by Spencer.[4] Burt's review of the results of statistical analysis (16), Vernon's description of test-results obtained from the Fighting Services during the war (17), and the accounts given by Cattell (12) and Eysenck (13) of the organization of personality, all support the view that factorial work reveals something like a hierarchical structure in the individual mind.[5]

post mortem the brains of children from 5 days to 13 years, finds that (i) the motor areas and sensory areas for vision, hearing, and kinæsthesis mature first ; (ii) the perceptual (visuopsychic, auditopsychic, etc.) mature a little later ; (iii) Broca's area later still ; and (iv) the so-called associative areas last of all ; not until the fourth year do those in the frontal lobe reach a stage comparable to that attained by the motor area as early as 11 months : (a convenient summary of this and similar studies, with references, will be found in Carmichael, L., *Manual of Child Psychology*, 1946, pp. 354f.).

[1] Hollingworth, H. L., *Psychology : Its Facts and Principles*, 1928. As can be seen from the arrangement of their ' Contents,' the conception, though not the name, is found in the text-books of many earlier writers, both structuralists and functionalists : Titchener's *Primer*, for example, deals successively with ' Sensation,' ' Perception,' ' Association of Ideas ' (leading to ' Memory and Imagination '), and finally ' Thought ' (which includes ' Concepts,' ' Judgement,' and ' Reasoning ') ; the corresponding affective states are interspersed. Much the same arrangement is found in Stout. Mellone and Drummond explicitly adopt the Spencerian terms and speak of ' mental development ' issuing in ' a *hierarchical* system in which different *levels* can be distinguished ' (*Elements of Psychology*, p. 71 : their italics). On the other hand, the earlier behaviourists tend to treat the progressive types of organization (by means of ' higher unit mechanisms ') as a result solely of learning, though presumably there must be some innate peculiarity of the neural structure which makes such an organization both possible and frequent.

[2] Tolman, E. C., *Purposive Behaviour in Animals and Men*, 1932, pp. 415f. Hull, C. L., *Principles of Behaviour*, 1943, pp. 42f.

[3] Allport, G. W., *Personality* (1938), pp. 139–42.

[4] Piaget, J., *The Psychology of Intelligence*, 1950, pp. 123, 150f. Piaget recognizes two main levels : (i) pre-symbolic (simple sensori-motor processes, perceptual processes, and mechanical association) and (ii) symbolic processes (including imaginal processes, processes of intuitive judgement, and processes of formal reasoning). " Each level is characterized by a new co-ordination of processes of the previous level." Within each main stage or level development proceeds by the differentiation of more specialized abilities out of more general.

[5] Burt writes : " The facts " (as summed up in his review of factorial work among children) " suggest a complex and systematic type of structure, and go far to confirm the notion that the mind is organized on a hierarchical basis." He prints a ' hierarchical scheme of mental levels,' with afferent and efferent aspects, similar to the pedigree diagrams he had used to illustrate the factorial analysis of the Binet Scale and other test batteries : ((15), pp. 103, 199). Vernon prints a similar ' inverted tree,' and observes that tables giving factorial results for adults tested in the Army and Navy " bring out a feature which appears highly characteristic of mental structure, namely, hierarchy." He adds that " the hierarchical theory . . . should be regarded as approximate only " ; and notes that, like himself, " Cattell regards the hierarchical picture of abilities as superior to the multiple factor " ((16), pp. 22, 24). Cattell asserts that " among the various terms used to describe classes of traits it is possible to recognize a hierarchy based on the width of reference of the trait," and believes it may be possible " to equate the three levels of factor, general, group, and specific, to three levels of trait breadth " ((12), pp. 131, 134, 143). Eysenck states that " a hierarchical view of personality structure, akin both to Burt's conception of factors and McDougall's theory of sentiments, would appear to account for the known facts better than any other." He too gives a pedigree diagram showing " four levels of organization corresponding closely to the four types of factor " ((13), pp. 16, 29, 30). On the orectic or conative side, the notion has a long history, for here the ' higher ' levels can readily be conceived as literally controlling or governing (ἀρχή) the lower : e.g., as Professor Duncan-Jones has recently pointed out, Bishop Butler's conception of ' human nature ' is essentially ' hierarchical ' (*Butler's Moral Philosophy*, 1952, pp. 44–8).

Except, however, for those statistical investigators who have followed either Burt, Spearman, or Thorndike, the supporters of the hierarchical theory seldom explain how precisely they use the term. In particular, the three alternative versions described above seem to be continually confused under the same general name. And neither those who accept the theory nor those who reject it appear to have attempted any direct attack upon the issues : acceptance or rejection seems to follow as an incidental consequence of some broader preconception. For these reasons, therefore, it seemed desirable to attempt a more direct study of the problem in and for itself.

II. ALTERNATIVE HYPOTHESES AND METHODOLOGICAL ASSUMPTIONS

Hypotheses. In the light of the foregoing review, the chief alternative hypotheses between which we have to decide may be set out as follows.

1. Starting with the most general question of all, the first hypothesis that calls for examination is the view that, on the cognitive side at any rate, mental characteristics can be classified in accordance with some type of hierarchical principle : that is, they can be arranged in some kind of progressive sequence, either branching, graded, or discrete.

2. If the evidence gives an affirmative answer to the question in its broadest form, we shall then have to ask what is the precise nature of this principle. As we have seen, there are three very different interpretations.

(*A*) Taking the term in its original and technical sense, the second and more specific hypothesis to be examined will be the hypothesis of ' subdivided factors,' i.e., the view that mental processes exhibit (i) a *generic* factor covering every type of cognitive process from the lowest to the highest, (ii) more *specific* factors distinguishing the ' lower ' processes from the ' higher ' (with possibly one or more intermediate groups), (iii) still more *specialized factors* into which these broader factors can be subclassified, and so on.

To this subdivided scheme there are two main alternatives : namely,

(*B*) that the so-called hierarchical arrangement involves simply a serial gradation, marked only by differences in complexity, such as could be described by a single general factor ; and

(*C*) that there is no general factor whatever, but merely a small number of group factors, with the successive levels differing in kind or quality rather than in degree.[1]

In analysing processes so elusive and involved as those of mental life, we may not perhaps succeed in finding any neat or clean-cut mode of classification, like those depicted in the tabular diagrams representing the various theories. What we must be prepared to discover will be an intricate and modifiable structure such as generally characterizes living organisms, not the rigid, regular, and symmetrical architecture characteristic of an artificial product. All we can expect, therefore, will be some sort of approximation to one scheme rather than another. Hence the issue will turn chiefly on the *degree* of approximation discernible in the facts observed ; and will consequently have to be decided by a statistical rather than a merely logical analysis.

Criticisms of the Procedure Proposed. During the last year or two there has been a good deal of criticism of the statistical procedures adopted for such problems as the present. The whole conception of ' factors ' has been questioned, particularly

[1] These are not the only conceivable hypotheses. There is, for example, the notion, suggested by several earlier writers, that the simplest and oldest capacities (such as sensory discrimination) are the most general, and that perceptual ability, and then general associative ability, are superadded (see below, p. 172). But to-day no writer to my knowledge supports such a scheme. And in any single research " we cannot deal with an exhaustive list of *all* conceivable hypotheses, and may reasonably limit ourselves to those that seem to have the greatest *a priori* probability " (Burt, *loc. cit. sup.*).

by writers with neurological [1] or psychiatric [2] interests. Their chief complaint is that "mathematical psychologists make no attempt to state or justify the assumptions which the use of such methods implies."

"Factorists," we are told, "take it calmly for granted that test performances can be 'analysed' into elementary abilities, and regard them as formed by adding separate quantities of *G*, *N*, *V*, and the like. Their approach is modelled on that of the chemist who discovers that caustic potash can be analysed into three simple elements of K (potassium), O (oxygen), and H (hydrogen), and conversely that by adding 16 grams of O and 1 gram of H to 39 grams of K he can obtain 16 + 1 + 39 = 56 grams of the compound, KOH. But it is absurd to suppose that the mind can be treated as the *sum* of atomistic elements: and even in chemistry the properties exhibited by a compound are seldom the sum of the properties possessed by its several ingredients."

Now all such arguments grossly misunderstand what the factorist attempts to do. Their widespread acceptance seems due to the fact that non-statistical psychologists are much more familiar with the idea of chemical analysis than with that of vector analysis; and consequently confuse the two. No doubt the earlier experimentalists were a little too fond of illustrating their methods by analogies drawn from chemical investigations. Sully,[3] for example, Spearman's predecessor at University College, starts his text-book by telling us that "the first and main method of the science [psychology] consists in the search for primitive elements, and therefore begins with an analysis into constituent *elements or factors*." In a footnote, however, he emphatically warns his readers that, unlike the chemist, the psychologist "attempts an ideal or logical separation only; he is not concerned with the actual separation of a concrete substance": for him "to analyse is simply to *classify*" (Sully's italics). Those who followed Sully adopted much the same language, but forgot the warning.[4]

It was partly to avoid this misunderstanding, as Burt has recorded, that he preferred to retain the term 'factor' and drop the term 'element.' For him 'factors' are not concrete parts or entities, but abstract 'principles of classification.' But unfortunately the word 'factor' has itself turned out to be highly ambiguous. As the *Oxford Dictionary* reminds us, it has two distinct senses, an abstract and a concrete: (i) "one of the components that make up a number or expression," and (ii) "a circumstance, fact, or influence contributing to a result." Burt, in using the term to describe the results of his particular form of analysis, plainly has in mind the former sense.[5] His analysis resembles, not the chemist's analysis of compounds into elements, but (as he repeatedly explains) the mathematician's resolution of 'directed quantities'—motions, forces, and the like—into rectangular components or dimensions.

Professor Zangwill, however, states that "in explaining how *the factorial analysis of ability* came to be developed, we cannot do better than quote Spearman's own account." [6] But in Spearman's earlier writings there is nothing which in any way resembles what is nowadays called factor analysis; and when later on he began regularly to use the term 'factor,' he interpreted it in a much more

[1] E.g., Zangwill, O. L. *Introduction to Modern Psychology*, 1950, esp. ch. VI, 'Intelligence and Mental Measurement.' A still fuller exposition of this standpoint is to be found in Chambers' article on 'Statistics in Psychology' (*Brit. J. Psych., Gen. Sect.*, XXXIII, pp. 189f.) from which Professor Zangwill quotes.

[2] E.g., Kirman, B., *This Matter of Mind*, 1951, esp. ch. III, 'Human Abilities,' and pp. 44f., on the methods of "Pearson and his followers." For a fuller reply to Kirman's criticisms I may refer to my review of his book in the last issue of this *Journal* (V, 1952, pp. 135–6).

[3] *The Human Mind: A Textbook of Psychology*, 1892, pp. 23f.

[4] Spearman, for example, in his first article ('General Intelligence Objectively Determined and Measured,' *Amer. J. Psychol.*, XV, 1904, pp. 201f.) is evidently thinking of chemical analysis rather than statistical analysis: he compares his experiments to the passing of a current through water to separate it into oxygen and hydrogen. 'General intelligence' is described as the 'common and essential element' in the 'sensory functions' and contrasted with the 'specific elements' with which it is supposed to combine.

[5] *Brit. J. Psych.*, III, 1909, p. 95: e.g., when Burt there refers to what he afterwards identified with 'intelligence' as the 'highest common factor,' the word factor is obviously used in a mathematical sense.

[6] *Loc. cit.*, p. 157. If the reader will turn to the passage from which Professor Zangwill is quoting (*Abilities of Man*, p. 73), he will find that "Spearman's own account" refers, not to the development of factor analysis, but to the development of his own hypothesis of "the universality of *g*," and of the "criterion" which he proposes for demonstrating the non-existence of any mental factors or faculties besides *g*.

concrete sense. Indeed, he actually quotes the second definition given above, and expressly adopts it. Hence Spearman's views can hardly be taken as typical.[1]

The idea that mental phenomena could be described in terms of abstract *Faktoren oder Komponenten* is to be found in several older writers, e.g., in Wundt's explanation of colour-mixture and in his account of his tridimensional theory of the feelings ; and Wundt's intention is perfectly clear. Since he speaks of the *Komponenten* and *Resultanten* as obtained by analysing or combining ' directions ' in a ' three-dimensional manifold,' he is manifestly thinking in terms of the ' resolution ' or ' composition ' of abstract vector-quantities, not of concrete causal agencies.[2] Wundt, however, does not use this ' mathematical model ' to verify his hypotheses, but only to explain them.

Burt's contribution was to adopt or adapt for psychological purposes the statistical technique described by Pearson for determining ' lines of closest fit ' ; and this in turn, as Pearson himself observes, was based on the traditional method of analysing what mathematical physicists had begun to call ' vector-quantities.'[3] Later writers, who speak of ' factor analysis '—Kelley, Hotelling, Holzinger, Thurstone—all use the term ' factor ' in this abstract sense. It retains the same meaning in Fisher's analysis of variance. Hotelling, however, expressly tries to avoid the ambiguity by substituting the more technical term ' component.' The application of the formulæ used in the composition and resolution of forces was perhaps most natural in the case of emotional tendencies, since such tendencies may readily be regarded (in McDougall's phrase) as mental ' forces.' But the same procedure may be applied equally well to the analysis of cognitive ' abilities,' if those abilities are measured in terms of the relative difficulty of the tasks which they are used to overcome.

Assumptions. What then are the assumptions involved in such procedures ? Here it is scarcely possible to offer a complete defence of factorial methods. But, in view of recent criticisms, it is essential to give some brief account of the assumptions I propose to adopt. Perhaps the following quotation will suffice.

" To express a set of complex objects, classifiable quantitatively under different qualitative categories, in a form amenable to mathematical treatment, (i) their different *quantities* must all be measurable in terms of a scale homogeneous for all the tests (e.g., the ' difficulty ' of the process tested), and (ii) their different *qualities* must be represented as different directions in multidimensional space. The *similarity* of any two quantities can readily be indicated by the similarity of the corresponding directions, measured, for example, by the smallness of the angle between them, or by the smallness of its cosine. We can then (iii) treat the observable quantities as ' resultants of *hypothetical components*,' the combination following the so-called parallelogram law. The relation between the observed test-correlations and the hypothetical test-factor correlations will then be given by the generalized equation for the cosine of a compound angle, which we can re-write in the form

$$r_{ij} = r_{ix}r_{jx} + r_{iy}r_{jy} + \ldots + r_{iz}r_{jz}. \qquad \text{(i)}$$

This formula therefore may be regarded as expressing an ' analysis ' of the *tests* i and j into the *factors* $x, y, \ldots z$. In most enquiries it will be preferable to choose factors that are independent of (i.e.,uncorrelated with) one another. In that case, to determine r_{ix}, etc., we may take

$$r_{ix} = \frac{\sum_j r_{ij}}{\sqrt{\sum_i \sum_j r_{ij}}}. \qquad \text{(ii)}$$

No doubt, as applied to any empirical phenomena, such a model, like most mathematical models, will be approximate only. Predictions based on it can claim no more accuracy than, say, predictions based on meteorological analysis. ' The wind bloweth where it listeth,' and at a speed which is by

[1] *Brit. J. Psych.*, XXIX, 1938, p. 184. The definition is given by Spearman in reply to Thomson's protest that in the mathematical sense a ' factor ' is *not* a concrete ' entity or organ,' but a mere abstraction. In Spearman's 1904 article the word ' factor ' is only used incidentally in reference to such " irrelevant influences " as age, sex, and practice ; it is never applied to ' general intelligence.' The name ' two-factor theory ' was adopted some ten years later by Spearman from De Sanctis. De Sanctis had declared that Spearman's ' unifocal hypothesis ' was equivalent to simplifying the theory of *three* kinds of factors (general, group, and specific) by discarding the group factors, on the ground that they were merely the old and discredited ' faculties ' renamed.

[2] *Grundzüge der Physiologischen Psychologie*, II, pp. 211, 298, 318, and Fig. 230.

[3] A vector-quantity may be defined as one which involves for its specification a statement of its direction as well as its magnitude. The term is due to Hamilton (1843) ; but its application in physics is due chiefly to Willard Gibbs (1891). A convenient introduction to the subject will be found in C. C. Macduffee's *Vectors and Matrices* (1943) : for the addition and multiplication of vectors see pp. 16, 17, and for the ' principal axis transformation ' (the mode of resolution proposed by Pearson) see p. 171. It should be noted that the ' sum ' of two vectors is their ' resultant ' as given by the familiar parallelogram law (p. 16).

no means uniform : but that does not prevent the meteorologist from analysing the general direction of a wind in terms of hypothetical components orthogonal to each other, or from expressing its effective speed as so many miles per hour." [1]

From this statement it will be seen that most of the criticisms raised by Zangwill, Chambers, and Kirman are really beside the point. *They are levelled against alleged assumptions which factorists do not as a matter of fact adopt.* Those critics who object to the assumption that factors are " added like so many grams of this or that elementary substance " should note that the process assumed is not scalar addition (as for substances that can be weighed) but vector addition (as for forces, or points referred to their co-ordinates). Those who object to the assumption that factors are " ultimate and indivisible entities similar to the phrenologist's faculties or organs " should also note that the statistician's factors are expressly declared to be logical abstractions—principles of classification rather than ' primary abilities ' or ' elementary traits ' ; the factorist does not claim a distinct and concrete existence for them any more than a meteorologist supposes that what he calls a north-east wind is formed by adding an actual north wind to an actual east wind. And, indeed, one of the consequences of the hierarchical theory is that the factors discovered are neither ultimate nor indivisible : " as each successive factor is extracted the initial group of traits is divided into semi-groups ; each of these in turn into demi-semi-groups ; and these into hemi-demi-semi-groups ; and so on." The popular analogy between mathematical analysis into factors and chemical analysis into atomic elements is thus highly misleading.

How then are we to justify a set of assumptions such as those formulated above ? To my mind, their chief vindication is to be found, not in any *a priori* speculations about the nature of the " human mind or personality," but rather in the *a posteriori* consequences to which those assumptions lead. In short, whether the procedure proposed is in fact practicable, and, if so, whether the results thereby achieved will possess any intelligible meaning are questions to be decided by actual trial.[2]

III. PLAN OF EXPERIMENT

(1) *Tests.* The majority of the tests used in the present investigation were specially adapted or devised for the purpose. As far as practicable, they were constructed so that each should depend on processes supposed to be distinctive of one cognitive level only. Naturally it is impossible to test perceptual processes without also evoking sensory processes, or associative and relational processes without also introducing perception. But in each case an endeavour was made to reduce the influence of the irrelevant accompaniments to a minimum. Several minor studies with a wide variety of test-materials and test-procedures were first undertaken with small groups of children, most of whom were personally known to the investigator ; and a tentative battery was thus evolved for preliminary trial under school conditions. The results of this pilot enquiry showed that several of the tests originally devised were unsatisfactory. These were discarded. Twenty were retained. With the split half method their reliability was found to vary from 0·70 to 0·90.

A brief description of the test material and procedure is appended. For fuller details the reader may refer to (18). Tests 1, 2, and 3 were administered individually, the rest in group form.

A. Sensori-Motor Level

Test 1. *Touch Discrimination.* The cutaneous threshold was measured in the usual way on the forearm by an æsthesiometer. The psychophysical procedure adopted was McDougall's method of serial groups.

Test 2. *Weight Discrimination.* 21 cans arranged in seven sets of four were loaded with shot to give different weights. In each set the ratio of the differences between the successive weights was diminished until they were just below the expected threshold, so as to provide tests of differing difficulty. The child was asked to arrange each set of four in order of heaviness. After each trial he

[1] This statement is taken, with some slight modification and abridgement, from Burt. For a fuller account, see his *Factors of the Mind*, 1940, pp. 115f.

[2] For an early and fuller defence of the use in psychology of the foregoing ' cosine law ' (eqn. (i)), see Garnett, J. C. M., *Brit. J. Psych.*, 1919, pp. 359f. It may be noted that this equation is adopted by Thurstone and described by him as the ' fundamental factor theorem ' (*Vectors of Mind*, 1935, p. 66, eqn. (37) and (39) ; cf. p. 92) ; and what he calls a centroid analysis is based on Burt's eqn. (ii) (*loc. cit.*, p. 94, eqn. (13)). Thurstone's identification of his factors with ' mental faculties ' (p. 53) or ' primary abilities ' (or ' traits,' p. 157) may perhaps have lent some colour to the criticisms quoted above.

was assigned 2 marks for a correct arrangement, and 1 mark when only two weights were correctly arranged.

Test 3. *Simple Reaction Time with Auditory Stimuli.* The stimulus was a sound-hammer ; and the subject had simply to press a key as soon as he heard the sound. The speed of reaction was measured with a chronoscope, and the average time taken for 10 trials.

Test 4. *Speed of Writing.* The child was required to write the figures 1 to 9 over and over again as rapidly as possible for the space of one minute. Three trials were used, separated by rest pauses of a minute ; and the average time taken for the three.

B. Perceptual Level

Test 5. *Counting.* On large sheets of paper, distributed to the children in class, a number of crosses were printed, arranged in two sets. In the first the crosses were arranged in rows, in the second in squares. The child was asked to count the number in each row or square. The total number counted in 2 min. 20 sec. was taken as the measure of his performance.

Test 6. *Checking Names and Numbers.* The subject was shown 20 pairs of numbers consisting of 4 or 5 digits, and then 19 pairs of names consisting of 6 or 7 letters. In each pair the second number or name was sometimes identical with the first, and sometimes differed by the alteration of one or two digits or letters. The child was required to decide as quickly as possible whether the numbers or names in each pair were exactly the same or not. The number correctly marked in 2 min. 30 sec. was taken as the measure of his performance.

Test 7. *Sorting Shapes.* Seven different shapes (simple geometrical figures), each labelled with a distinctive letter, were drawn at the top of a sheet of paper to serve as a key. Below were 56 drawings consisting of the same shapes arranged in random order. The child was required to mark each with the appropriate letter.

Test 8. *Perception of Parts.* Five geometrical figures, each labelled with a distinctive number, were shown at the top of a sheet ; and parts of these figures were presented below. The child was asked to write against each a number showing to which of the whole figures the several parts belonged.

Test 9. *Perception of Patterns.* The child was given a sheet containing on the left-hand side a number of crosses, each set forming a distinctive pattern. On the right-hand side there were groups of crosses arranged in sets ; and the child was asked to draw a line round the crosses in each set on the right, so as to reproduce the patterns on the left.

C. Associative Level

Test 10. *Memory for Numbers.* Series of numbers, consisting of 3, 4, 5, 6, 7, and 8 digits, were read aloud ; and the child was asked to write down each set in the correct order directly it had been read.

Test 11. *Memory for Shapes.* Twelve shapes were set before the child, and he was given 2 min. 30 sec. to study them. They were then mixed with a number of slightly different shapes ; and he was asked to mark those he recognized.

Test 12. *Visual Imagery.* This test contained 60 items. In each a simple diagram was shown, representing a card of some particular shape with its face upwards : at the side other drawings of the same card were shown, representing it in various positions, sometimes face upwards, sometimes turned over with its face downwards. The child was asked to indicate those that showed the card lying face downwards.

Test 13. *Association.* This consisted of two parts : in the first, a single stimulus word was given, and the child was required to write down all the words that the given word suggested ; in the second, he was asked to write down all the words he could think of belonging to a definite group or topic (e.g., flowers, fruits, things that can be seen in school, things that can be seen in the street, etc.).

D. Relational Level

Test 14. *Synonyms and Opposites.* Lists of words, arranged in pairs, were set out on a sheet of paper, each pair consisting either of two synonyms or of two opposites. The child was required to underline the words ' Same,' ' Opposite,' or ' Do not know,' printed against each pair.

Test 15. *Verbal Analogies.* This was the well-known analogies test introduced by Burt. It consisted of 20 sets of 3 words, constructed specially for this experiment. The child had to discover the relation between the first two, and then write down a word standing in the same relation to the third as the second did to the first.

Test 16. *Non-Verbal Analogies.* In this version of the test the terms of the relations were indicated, not by words, but by pictures. For the fourth term four alternative pictures were shown ; and the child had to mark the correct picture with a cross.

Test 17. *Completion.* A story was typed out with crucial words omitted : (it will be found in *Mental and Scholastic Tests*, pp. 233–4). The child was told to write a suitable word in each blank space.

Test 18. *Syllogisms.* Each problem consisted of two premisses, with the conclusion to be deduced from them indicated in skeleton form only, the crucial word being omitted. The child had

to insert the correct word (e.g., " A is larger than B ; C is smaller than B ; therefore A is . . . than C "). This is perhaps the most representative of Burt's ' reasoning tests ' (cf. (6), p. 102).

Test 19. *Number Series.* Sets containing 4 numbers in each were printed on a sheet, the numbers being so chosen as to indicate a sequence obeying some simple rule. The child had to write down the next two numbers in the sequence.

Test 20. *General Intelligence.* Finally, Form A from the Otis Advanced Group Test was used to obtain an assessment of each child's intelligence.

(2) *Subjects.* One hundred and sixty-six boys, aged 10.0 to 11.0 years, were tested. They consisted of all the pupils within the age limits specified who were attending four London schools.[1] The schools were carefully chosen so that the children to be tested could be regarded as typical of the average elementary school population. Although attending different schools, they came from similar homes and lived in similar types of neighbourhood. To verify the assumption that the recruitment of the schools was much the same, an analysis of variance was carried out. No significant differences were found in either the means or the variances ; and the frequency distribution obtained from the various tests (each checked by the chi-squared criterion) conformed, as closely as could be expected with a sample of this size, to a normal distribution, Accordingly, in what follows the results will be treated as coming from a single group.

IV. ANALYSIS OF RESULTS

Methods. Correlations were calculated between the results for all the tests by the usual product-moment formula. The coefficients so obtained are set out in Table I. In this and the following tables the tests have been rearranged in order according to the total correlation obtained for each. Of the 190 coefficients all are positive, and all except three are statistically significant.

In order to ascertain what type of factor-pattern will give the best fit to the correlations obtained, it will be necessary to apply in turn all the main methods of factor analysis. If the investigation were confined to the results of one method only, those who favour some alternative procedure might be tempted to wonder whether it might not have produced a very different conclusion. The methods employed were the following ; all depend, at some stage of the calculation, on the two fundamental equations given above (eqn. (i) and (ii), p. 160) :—

A. Undivided Factors :
1. For Bipolar Factors : Simple summation.
2. For Group Factors : (*a*) Analysis into non-overlapping group factors ; (*b*) Arithmetical rotation to obtain overlapping group factors ; (*c*) Graphical rotation to obtain simple structure ; (*d*) Bifactor method ; (*e*) Triangular method.

B. Subdivided Factors :
1. For Bipolar Factors : Simple summation with subdivided factors.
2. For Group Factors : (*a*) Direct group factor analysis with subdivided factors ; (*b*) Arithmetical rotation to obtain a complete factor matrix.

A. Undivided Factors

1. *Bipolar Analysis.* With a correlation table such as that given on p. 164 the natural starting point is a straightforward analysis by means of simple summation. The formulæ and iterative procedure described by Burt were used.[2] The saturations so obtained are set out in Table II A. In all, five factors were extracted, but only three proved to be statistically significant.

[1] I am specially indebted to the headmasters for their willingness to allow me to test the pupils in their charge and for their generous co-operation throughout the whole experiment.
[2] See Burt, *Factors of the Mind* (1940), Appendix, pp. 462f. The iterative procedure is required to obtain satisfactory values for the self-correlations or ' communalities.'

TABLE I. CORRELATIONS BETWEEN TESTS

Tests	20	14	19	16	17	18	15	13	11	12	10	8	7	9	5	6	4	3	2	1
20. Mental Age	(·854)	·842	·813	·805	·826	·820	·838	·586	·556	·570	·552	·347	·383	·339	·296	·256	·229	·196	·156	·217
14. Syn. and Opposites	·842	(·849)	·795	·785	·842	·791	·791	·624	·514	·548	·456	·347	·309	·239	·365	·383	·248	·216	·228	·244
19. Number Series	·813	·795	(·773)	·815	·757	·773	·769	·567	·531	·548	·556	·349	·346	·298	·325	·334	·256	·228	·215	·207
16. Non-Vbl. Analogies	·805	·785	·815	(·800)	·811	·768	·804	·537	·552	·555	·519	·353	·350	·312	·276	·341	·198	·211	·200	·211
17. Completion	·826	·842	·757	·811	(·829)	·770	·817	·583	·520	·522	·455	·292	·313	·285	·268	·345	·197	·205	·209	·175
18. Syllogisms	·820	·791	·773	·768	·770	(·788)	·786	·559	·465	·506	·527	·350	·343	·294	·323	·389	·201	·177	·218	·185
15. Vbl. Analogies	·838	·791	·769	·804	·817	·786	(·805)	·563	·520	·494	·459	·285	·294	·304	·294	·238	·195	·236	·164	·216
13. Association	·586	·624	·567	·537	·583	·559	·563	(·590)	·560	·581	·563	·331	·332	·268	·325	·271	·213	·179	·197	·188
11. Mem. Shapes	·556	·514	·531	·552	·520	·465	·520	·560	(·635)	·665	·570	·332	·312	·363	·317	·258	·234	·155	·201	·191
12. Imagery	·570	·548	·548	·555	·522	·506	·494	·581	·665	(·631)	·542	·273	·395	·330	·255	·283	·176	·207	·229	·225
10. Mem. Numbers	·552	·456	·556	·519	·455	·527	·459	·563	·570	·542	(·546)	·320	·330	·278	·319	·195	·236	·249	·185	·202
8. Perception Tests	·347	·347	·349	·353	·292	·350	·285	·331	·332	·273	·320	(·310)	·323	·325	·308	·306	·215	·199	·305	·206
7. Sorting Shapes	·383	·309	·346	·350	·313	·343	·294	·332	·312	·395	·330	·323	(·411)	·380	·349	·349	·166	·228	·174	·047
9. Perception Patterns	·339	·239	·298	·312	·285	·294	·304	·268	·363	·330	·278	·325	·380	(·440)	·351	·303	·175	·219	·152	·128
5. Counting	·296	·365	·325	·276	·268	·323	·294	·325	·317	·255	·319	·308	·349	·351	(·448)	·279	·278	·174	·089	·225
6. Checking	·256	·383	·334	·341	·345	·389	·238	·271	·258	·283	·195	·306	·349	·303	·279	(·313)	·229	·168	·180	·239
4. Writing Speed	·229	·248	·256	·198	·197	·201	·195	·213	·234	·176	·236	·215	·166	·175	·278	·229	(·284)	·298	·268	·272
3. Reaction-time	·196	·216	·228	·211	·205	·177	·236	·179	·155	·207	·249	·199	·228	·219	·174	·168	·298	(·311)	·287	·278
2. Weight Discrimination	·156	·228	·215	·200	·209	·218	·164	·197	·201	·229	·185	·305	·174	·152	·089	·180	·268	·287	(·276)	·297
1. Touch Discrimination	·217	·244	·207	·211	·175	·185	·216	·188	·191	·225	·202	·206	·047	·128	·225	·239	·272	·278	·297	(·320)

The first factor contributes 41·5 per cent. to the total variance. For this all the tests have positive saturations. Moreover, it will be seen that, on the whole, the size of the saturations corresponds with the complexity and the level of the various processes tested : tests intended to measure processes on each higher level have larger saturations than those obtained for tests intended to measure processes on the levels below. The highest saturation of all is obtained by Mental Age as assessed by the Otis Intelligence Test ; the factor may therefore be regarded as a factor of general cognitive ability or ' intelligence.' Thus, at the very outset we find that the results of the experiment are fully consistent with the first of our hypotheses. That, of course, does not so far imply that they must be inconsistent with any particular supplementary explanation.

The second factor is bipolar, and contributes 6·7 per cent. to the total variance. A factor of this size is unquestionably significant ; and we thus have clear evidence (if fresh evidence were required) showing that a single general factor alone is not sufficient to account for the facts. What is more instructive is the nature of the factor. It classifies the entire set of tests into two main groups : the first or positive half includes all the tests for sensori-motor and perceptual processes ; the second or negative half includes all the tests for associative and relational processes, whether their material is verbal, numerical, or concrete. The distinction here drawn resembles that mentioned by Burt in his survey of the results of factor analysis, where he contrasts (i) the *practical* type of person in whom sensory and perceptual experience predominates, together with the direct motor reactions to which they naturally lead, and (ii) the more *intellectual* type of person, in whom the secondary associative processes predominate, bringing with them memories, images, and ideas, and leading to reasoning and reflection rather than to immediate action. " Popular opinion," he writes, " opposes the practical or observant type to the thoughtful or reflective. . . . The contrast seems preserved by so cautious a psychologist as Professor Stout, who distinguishes the perceptual from the conceptual or ideational levels of the mind. This broad classification corresponds to the traditional distinction between ' lower ' and ' higher ' mental processes, and the phrenological division of ' intellect ' into ' perceptive ' and ' reflective ' processes." [1]

The third factor contributes 3·7 per cent. to the total variance. It apparently classifies each of the broader groups into two subgroups : the first group is divided into (i*a*) the simple sensory and motor tests and (i*b*) the more complex perceptual tests ; the second into (ii*a*) the tests of the more mechanical associative processes and (ii*b*) the tests of relational processes.

There is one exception. Test 6 (' Checking '), which we should expect to be classified with the tests in the perceptual group, has a small negative saturation and thus becomes allocated to the sensori-motor group. This perhaps is not surprising. Test 6 proved to be the simplest of all the perceptual tests, and little more than a complex reaction test. Moreover, there are five perceptual tests, some with large saturations, and only four sensori-motor tests, all with small saturations. In such a case, the requirement that the saturations in each half should sum to zero may easily change the sign of the smallest saturation coefficient in the overweighted group.

The precise nature of this third factor is not easy to interpret from the saturations alone. As we have seen, it introduces a subclassification within each of the two broader groups into which the tests are separated by factor II. But the basis of this subclassification can hardly be considered the same within each group. The differences are not merely differences in complexity, like those which seem to be measured by the saturations for the first factor : they are differences in the kind of process rather than in degree. This is borne out by introspective considerations. Perception, for example, as psychologists are fond of pointing out, is very much more than the simultaneous experience of a complex group of sensations as distinct from a single simple sensation ; association and the use of imagery are qualitatively different from either sensation or perception ; and the explicit awareness and use of relations are qualitatively different from the reliance on mere mechanical and unconscious association.

[1] (10), pp. 10, 31, 33. He remarks that the so-called ' intellectual ' or reflective ability is contrasted sometimes with the ' perceptual ' and sometimes with the ' practical ' ; and he adds that he was originally inclined to believe that the fundamental distinction represented a distinction of ' aspects ' rather than of ' levels.' However, as he points out, two different principles of classification may here be involved ; and the distinctions may be as much temperamental as cognitive—related to the distinction drawn by Binet between ' objective ' and ' subjective ' types and by Jung between what he preferred to call ' extraverted ' and ' introverted ' types (cf. *ibid.*, p. 36). For the phrenologist's analysis of ' intellect ' into ' perceptive ' and ' reflective powers,' see Hollander, B., *Scientific Phrenology* (1905), Synoptic Table on p. 59.

TABLE II. SATURATIO

Methods	A. Bipolar					B. Group : Non-Overlap			
Tests	I	II	III	IV	V	I	II	III	IV
20. Mental Age	·850	+·334	+·081	+·114	−·014	·731	·569	—	—
14. Syn. and Opposites	·845	+·261	+·221	+·075	+·111	·736	·527	—	—
19. Number Series	·832	+·249	+·122	+·058	+·030	·742	·475	—	—
16. Non-Vbl. Analogies	·828	+·285	+·136	+·104	−·065	·720	·524	—	—
17. Completion	·813	+·328	+·205	+·097	−·093	·681	·591	—	—
18. Syllogisms	·814	+·253	+·160	+·169	+·090	·708	·510	—	—
15. Vbl. Analogies	·801	+·335	+·185	+·078	−·105	·665	·615	—	—
13. Association	·699	+·146	−·212	−·115	+·138	·737	—	·165	—
11. Mem. Shapes	·686	+·102	−·350	−·175	−·037	·702	—	·397	—
12. Imagery	·692	+·107	−·285	−·206	−·129	·713	—	·346	—
10. Mem Numbers	·654	+·072	−·262	−·197	+·074	·681	—	·235	—
8. Perception Tests	·493	−·249	−·044	+·052	+·001	·495	—	—	·249
7. Sorting Shapes	·498	−·215	−·219	+·229	−·129	·475	—	—	·412
9. Perception Patterns	·469	−·276	−·217	+·227	−·213	·438	—	—	·424
5. Counting	·476	−·278	−·165	+·184	+·287	·454	—	—	·326
6. Checking	·459	−·250	+·026	+·172	+·100	·452	—	—	·281
4. Writing Speed	·371	−·313	+·125	−·169	+·066	·341	—	—	—
3. Reaction-time	·358	−·321	+·146	−·190	−·146	·320	—	—	—
2. Weight Discrimination	·344	−·292	+·138	−·232	−·023	·305	—	—	—
1. Touch Discrimination	·346	−·278	+·210	−·273	+·057	·306	—	—	—
Contribution to Variance (per cent.)	41·5	6·7	3·7	2·8	1·4	35·2	10·4	1·8	3·0

To gain further light on the nature of this factor it may be helpful to examine the residuals on which it is based. If it was the last factor of all, or the last of the factors based on residual correlations of a significant size, we should not only expect that the residual correlations between the two contrasted halves within the same broad group would be *negative* (i.e., that the sensori-motor tests would be correlated negatively with the perceptual and the associative with the relational) ; we should also expect that the residual correlations between the two subsections, having similar signs, but coming from different groups, would be *positively* correlated (e.g., that sensori-motor tests would be correlated positively with relational, and perceptual with associative). But if they existed, these positive correlations would certainly be rather puzzling. However, when we turn to the actual residuals we find that no such positive correlations are to be discovered. The submatrices of cross correlations consist almost exclusively of near-zero or non-significant residuals. Further, the fourth factor, when extracted, proves to have a pattern of signs that largely annuls the theoretical cross correlations which result from the table reconstructed from them. There is one obvious explanation of all these anomalies. The idea of a second bipolar factor, which, like our factor III, covers the *whole* series of tests, is quite unsuited to the data. This second bipolar factor, therefore, which is forced to divide the whole set of tests into four co-ordinate groups, must be replaced by *two* bipolar factors, each extending over one-half of the tests only. In short, we really need to carry out a fresh analysis using ' subdivided factors ' instead.[1]

2. *Non-overlapping Group Factors.* Before we attempt this alternative type of analysis, let us first enquire what would be the result of substituting group factors (with positive saturations only) for bipolar factors (with both positive and negative

[1] As has been pointed out in previous papers (e.g. (15)), when the sign pattern of any particular bipolar factor, instead of introducing new subdivisions into the factor-matrix, simply repeats the divisions of its predecessor, with a reversal of just one-half of the signs, that is a fairly safe cue for the use of ' subdivided ' factors.

DIVIDED FACTORS

C. Group : Overlapping				D. Simple Structure					E. Bifactor				
II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
·555	−·056	·125	−·036	·012	·553	·516	·217	·465	·750	·544	—	—	
·485	−·207	·141	·083	·074	·383	·564	·231	·570	·733	·529	—	—	—
·478	−·084	·118	·052	·093	·467	·503	·220	·495	·736	·482	—	—	—
·539	−·040	·186	·019	·045	·512	·554	·230	·420	·711	·537	—	—	—
·599	−·068	·216	·048	·020	·491	·624	·194	·401	·672	·600	—	—	—
·450	−·178	·172	−·020	·004	·409	·507	·290	·530	·699	·514	—	—	—
·602	−·044	·196	·046	·025	·504	·609	·172	·388	·657	·615	—	—	—
·251	·077	−·200	·011	·225	·508	·116	·106	·504	·709	—	·222	—	—
·247	·309	−·214	·008	·304	·644	·046	·085	·345	·689	—	·396	—	—
·306	·347	−·155	·072	·330	·652	·133	·062	·276	·699	—	·363	—	—
·196	·184	−·228	·069	·317	·514	·058	·074	·416	·711	—	·236	—	—
−·072	·114	·157	·090	·290	·206	·102	·376	·178	·509	—	—	·232	—
−·057	·266	·220	−·143	·154	·386	·016	·483	·077	·477	—	—	·413	—
−·084	·338	·276	−·118	·189	·383	·016	·507	−·029	·446	—	—	·416	—
−·250	−·055	·042	−·091	·185	·133	−·106	·481	·392	·464	—	—	·309	—
−·116	·001	·213	·024	·186	·103	·110	·407	·336	·456	—	—	·273	—
−·124	·021	·091	·346	·436	−·013	·148	·210	·167	·331	—	—	—	·403
−·049	·174	·192	·390	·466	·058	·227	·195	−·014	·326	—	—	—	·436
−·071	·088	·093	·396	·464	·008	·180	·143	·090	·311	—	—	—	·453
−·068	−·007	·066	·456	·479	−·061	·218	·107	·164	·309	—	—	—	·430
11·7	2·9	3·1	3·1	7·2	16·7	11·9	7·6	12·7	35·0	10·5	2·0	2·8	3·7

saturations). The believers in group factors might hope that this procedure would serve to abolish the anomalies just mentioned. To obtain a group factor-matrix with positive saturations throughout, Burt's direct method of group factor analysis was adopted first of all ; this affords the simplest and surest way of verifying the presence of group factors, though it does not always indicate their real extent or range. In accordance with the routine instructions, the allocation of tests to the several groups was objectively determined by reference to the sign-pattern of the bipolar factors as obtained in the preliminary summational analysis. Since in the fourth factor the saturation for Test 6 (' Checking ')—which assigns it to the perceptual group—was much larger than the saturation in the third factor—which assigns it to the sensori-motor group—the test was classified with the former. The ordinary formula for four non-overlapping group factors was then applied.[1]

The results are set out in Table II B. As usual, for the first or basic factor the saturations prove to be somewhat lower than those furnished by the bipolar method. But the proportionate values are much the same. The separation between the values obtained for the first two levels is no longer quite so distinct ; indeed, one or two tests belonging to the relational level receive slightly lower saturations than those belonging to the associative level.

The saturations for the four supplementary factors confirm the presence of group factors distinguishing the four main levels. It will be noted that the factor for the relational tests here obtains the largest saturations : this is a little unusual, since the relational factor is usually one of the hardest to distinguish from general intelligence ; in this case perhaps the result is due to the rather large number of tests of a relational type which were included in the battery. The saturations for the associative or reproductive level are smaller than might be

[1] Burt, C., *Factors of the Mind*, p. 482, and ' Group Factor Analysis,' this *Journal*, III, 1950, pp. 50f.

anticipated, since in previous investigations memory and learning have usually provided well-marked group factors. On the other hand, the saturations for tests on the lowest and simplest level are larger than those commonly obtained.

When we turn to the residuals, however, it becomes clear that a set of non-overlapping factors does not by itself provide a very satisfactory fit to the table of correlations actually observed. It is evident that there must be either a good deal of overlap between the group factors indicated, or else one or two additional factors each broad enough to cover at least two levels.

3. *Overlapping Group Factors.* On these grounds it seemed essential to proceed to an arithmetical rotation in order to secure a more complete factor-matrix allowing for a possible overlap between the group factors. Accordingly, an orthogonal rotation matrix was calculated by the method described by Burt.[1] Since the analysis by non-overlapping group factors involves the assumption of five factors in all, a 5×5 rotation matrix was constructed and applied to the five columns originally obtained by the method of simple summation (the ' bipolar analysis '), in other words, to the first five columns in Table II.

The results are shown in Table II C. There are several rather unexpected features. The basic factor is but little altered. But the first of the group factors now extends over all tests of the two higher levels ; the second overlaps the second and the third levels ; and the third overlaps the first and third ; the last remains, as before, virtually confined to the tests of the simple sensori-motor level. Further, there are a number of rather large negative saturations. A detailed study of the aberrant figures will be suggestive : they point to the possibility, not only of one or two very broad factors, but also of certain very narrow group factors, arising out of similarities in the material content or the formal procedure used for certain of the tests. However, the figures themselves are too small to deserve discussion at any length.

4. *Simple Structure.* At this point the critic may object that it is quite conceivable that, with a different mode of rotation, four overlapping group factors might by themselves be sufficient to account for the observed correlations. In that case the general factor could be dispensed with. An attempt was therefore made to secure a factor-pattern conforming with Thurstone's requirements for ' simple structure.' [2]

A graphical rotation was adopted, based on his working procedure. Since Thurstone's formula for his centroid analysis is identical with Burt's original formula for simple summation, it was assumed that the factor-matrix giving general and bipolar factors (Table II A) would furnish an appropriate starting point. The factors were rotated in pairs. Some difficulty, however, was experienced owing to the fact that, at successive stages in the rotation, the computer has to make a choice between two or more plausible alternatives in selecting his pairs and in fixing the axes. Thurstone and his followers frankly base their decisions on the psychological meaning of the factors ultimately emerging. But, in the present case at any rate, this would manifestly beg the very question which it was hoped to prove. Accordingly, so far as possible, a ' blind rotation ' was attempted.

[1] The working method is quite simple, and is described in ' Group Factor Analysis,' *loc. cit. sup.* As regards the reasons for preferring an arithmetical to a graphical rotation, see Barakat, M. K., this *Journal*, IV, pp. 145–6. For the rotation matrices used in the analyses, the reader may refer to the tables in my thesis (18).

The routine arithmetical procedure assumes the use of a *square* rotation matrix ; and that in turn generally entails the inclusion of small bipolar factors that are not fully significant. This method would appear to be perfectly justifiable when these supplementary factors, though not fully significant, are more likely to be significant than not. But when (as here) one of the later factors is more likely to be the effect of chance than of any assignable cause, we ought perhaps to treat it as representing in the main a factor of correlated error. In such a case it might seem better to exclude it, since its tendency would be to distort rather than to improve the group factor-matrix obtained by rotation. However, first of all it was considered desirable to follow the customary procedure, in order to ascertain how far this expectation was actually borne out by the results.

[2] Thurstone, L. L., *Multiple Factor Analysis*, 1947, pp. 319f.

The results are shown in Table II D. With one or two minor exceptions, it was found possible to suppress almost entirely the negative factor-saturations. It was not possible to suppress the general factor altogether. As will be seen from the table, what is there called factor I still has positive saturations for every test, although the figures for the tests belonging to the highest mental level are greatly reduced in size. This reduction was only rendered possible at the cost of introducing a second extremely broad factor into the second column : here again all the saturations, except two, are positive, though with this factor it is the saturations for the tests on the lowest level that are reduced almost to zero. The final outcome is that the overlapping of the various factors becomes so wide and erratic that the distinction between the various levels is greatly obscured.[1]

Such a result, however, can scarcely be accepted as disproving the original hypothesis with which we are concerned.[2] For one thing, if we had adhered to Thurstone's own suggestion, and relied for guidance on the psychological interpretation of the factors to be reached, it would have been quite easy by continued rotations to secure a plausible fit with four narrow factors and one nearly general factor, and easier still to have reached a rotated factor-matrix similar to that given in Table II C, provided the investigator knows what to aim at. But, even were we willing to adopt this procedure, it seems evident that the outcome would not have been as good as that obtained with the more direct and speedier method of arithmetical rotation, starting with the bipolar-matrix as a guide. Moreover, the calculations for an arithmetical rotation are automatic, and therefore, unlike the graphical, quite uninfluenced by the computer's personal preconceptions : it therefore provides a far more objective verification.

5. *Bifactor Method.* As a check on the foregoing results, it was decided to carry out a bifactor analysis by means of Holzinger's procedure.[3] Here the allocation of the tests to the several groups is determined, not by reference to the sign-pattern of the bipolar or centroid analysis, but by the cumulative calculation of ' coefficients of belonging.'

The results of the analysis are shown in Table II E. The allocations and the saturations resemble those found with Burt's non-overlapping group factor method. But there are one or two minor differences. For example, as with the original summational analysis, the highest saturation for the general or basic factor is now obtained by the Otis test of intelligence. The saturations for the first factor are once again reduced in size. But some are altered far more than others. The three tests involving verbal reasoning, for instance, show saturations for the basic factor even smaller than many of the tests of association—a change possibly due to the fact that the former tests now obtain large saturations for the relational group factor.

[1] This kind of result seems rather apt to emerge when Thurstone's technique is used. For example, in Sutherland's ' orthogonal rotation ' (this *Journal*, III, Table III, p. 23) both factor I and factor V have positive saturations throughout ; with the former none of the figures, and with the latter only 4 out of the 21 figures, approach zero. (Sutherland says that " the guiding principle was rotation of axes to positions which could be interpreted psychologically." It must, however, be confessed that an interpretation which applies the label ' verbal ' to a factor in which three of the four highest saturations are *numerical* seems a little surprising.) Renshaw, who also uses Thurstone's methods for the same battery, gets two different factors with positive saturations throughout (this *Journal*, IV, Table V, p. 12). Here the lowest is 0·344. But to ' interpret ' two such factors, and give each a distinctive label, on the basis of a few small differences between pairs of saturations obtained in this arbitrary way, seems extremely precarious. He calls the first ' *g* ' and the second ' schooling ' : but he might almost equally well call the first ' schooling ' and the second ' *g*.' Surely such subjective procedures cannot possibly be accepted as providing objective verifications of the hypothesis to be tested—namely, the existence of separate factors for the acquired results of schooling and the innate effects of *g*. The conclusion only seems plausible because there is already plenty of evidence indicating the operation of these two distinct causal agencies.

[2] A similar point is made by Guttman in his recent article (' Multiple Group Methods for Common-Factor Analysis.' *Psychometrika*, XVII, 1952, p. 215). Commenting on Thurstone's method of rotating centroid factors, he observes : " *A posteriori* hypotheses, made *after* the inspection of the data, may be subject to all the uncertainties and controversies which beset any *a posteriori* theory." The aim of factor analysis, he adds, should be to use the data to test a hypothesis formulated *a priori*, i.e., before the commencement of the experiments. This is fully in accord with the principle quoted from Burt, p. 160 above.

[3] Holzinger, K. J., and Harman, H. H., *Factor Analysis*, 1941, pp. 111f.

TABLE III. SATURATIO

Methods	A. Bipolar				B. Group : Unrotated							
Tests	I	II	III	IV	I	II	III	IV	V	VI	VII	VII
20. Mental Age	·852	+·330	+·146	—	·702	·467	·389	—	—	—	—	—
14. Syn. and Opposites	·847	+·255	+·207	—	·676	·419	·434	—	—	—	—	—
19. Number Series	·835	+·246	+·131	—	·670	·479	·317	—	—	—	—	—
16. Non-Vbl. Analogies	·829	+·281	+·160	—	·642	·505	·359	—	—	—	—	—
17. Completion	·814	+·325	+·206	—	·653	·498	·368	—	—	—	—	—
18. Syllogisms	·813	+·243	+·197	—	·685	·330	·438	—	—	—	—	—
15. Vbl. Analogies	·802	+·333	+·218	—	·583	·517	·441	—	—	—	—	—
13. Association	·698	+·137	−·236	—	·604	·383	—	·244	—	—	—	—
11. Mem. Shapes	·685	+·092	−·392	—	·622	·248	—	·449	—	—	—	—
12. Imagery	·689	+·093	−·344	—	·625	·270	—	·431	—	—	—	—
10. Mem. Numbers	·652	+·059	−·293	—	·578	·305	—	·302	—	—	—	—
8. Perception Tests	·495	−·266	—	−·085	·523	—	—	—	·211	·128	—	—
7. Sorting Shapes	·492	−·201	—	−·320	·453	—	—	—	·159	·412	—	·17
9. Perception Patterns	·460	−·254	—	−·276	·477	—	—	—	·125	·302	—	—
5. Counting	·464	−·246	—	−·178	·441	—	—	—	·182	·287	—	·20
6. Checking	·457	−·250	—	−·122	·470	—	—	—	·208	·177	—	—
4. Writing Speed	·369	−·316	—	+·201	·333	—	—	—	·223	—	·320	·18
3. Reaction-time	·354	−·314	—	+·240	·314	—	—	—	·271	—	·342	·11
2. Weight Discrimination	·338	−·285	—	+·255	·249	—	—	—	·303	—	·365	—
1. Touch Discrimination	·339	−·262	—	+·283	·253	—	—	—	·191	—	·435	—
Contribution to Variance (per cent.)	41·4	6·4	3·3	2·8	29·7	9·4	5·4	2·7	2·1	1·8	2·7	0·6

B. Subdivided Factors

1. *Bipolar Analysis.* In applying the method of simple summation, it was noted that after the general and first bipolar factors had been extracted, the significant residuals were (with scarcely any exception) confined to two of the four quadrants into which the correlation table can be quartered, namely, to the two square submatrices astride the diagonal which contain respectively (*a*) the inter-correlations between tests of the relational and associative levels and (*b*) the inter-correlations between tests of the perceptual and sensori-motor levels : the residual figures in the two rectangular submatrices away from the diagonal were practically all non-significant : (cf. p. 166 above). For this reason and others already mentioned, a fresh analysis by simple summation was undertaken by the procedure required for ' subdivided bipolars.' [1]

The first two factors are left unaltered. The remainder of the analysis is then completed by taking the two submatrices that contain significant residuals and factorizing each submatrix separately and independently : except for this restriction the method of computation is precisely the same as that applied more usually to the correlation table as a whole. As a result, instead of a single bipolar factor entering into all the tests (like factor III in Table II A) two independent bipolar factors are obtained, each limited to about half the tests only.

[1] For a fuller description of the bipolar use of this procedure, and its justification, see Burt, C., ' Subdivided Factors,' this *Journal*, II (1949), p. 53, and Table II B ; in that article Table IV (*loc. cit.*, p. 56) gives a factor-matrix of subdivided group factors—seven factors in all—obtained for physical measurements, which seems almost exactly comparable with that obtained here for mental measurements.

BDIVIDED FACTORS

C. Group : Rotated

I	II	III	IV	V	VI	VII	VIII	IX
702	·431	·386	·049	—·031	·013	—·061	·133	·114
573	·412	·420	—·028	·034	—·015	·053	·121	·121
582	·395	·359	·034	·021	·005	·020	·117	·129
579	·406	·490	·009	—·002	·011	—·013	·139	·132
539	·428	·434	—·022	—·013	—·025	·001	·148	·124
661	·381	·397	—·042	·027	·049	—·025	·145	·098
630	·432	·419	—·002	—·021	—·034	·004	·139	·128
626	·303	·071	·279	—·007	·006	—·015	·003	·039
644	·222	—·034	·378	—·005	·021	·017	—·041	·042
636	·279	·005	·357	·002	—·014	·017	—·029	·051
598	·248	—·011	·329	·019	—·005	·033	—·036	·054
481	—·028	·017	·012	·193	·177	·109	·005	·033
517	·010	—·013	·036	·137	·298	—·091	—·002	—·069
502	—·031	·040	·023	·164	·317	—·068	·017	—·082
495	—·009	—·022	·014	·178	·283	—·014	—·003	—·071
452	·013	—·050	—·058	·186	·236	·046	·040	—·046
345	—·004	—·005	—·006	·244	·036	·328	—·026	—·058
329	—·011	—·007	—·008	·250	·032	·351	—·029	·046
308	·005	—·014	·017	·229	—·009	·360	—·036	—·048
294	·025	·018	—·003	·238	—·051	·412	—·028	—·037
51·5	7·4	6·1	2·3	1·9	1·8	2·8	1·5	1·4

The factor-saturations reached in this way are shown in Table III A. They give a distinctly better fit than is obtained with the matrix of undivided bipolars, and the residuals are negligible. The new factor-pattern thus demonstrates the presence of two broad group factors (as well as others), and avoids the awkward assumption that the subdivision of each broad group into two distinct levels is based on the same classificatory principle throughout (see above p. 165).

2. *Group Factors.* To obtain a factor-matrix of positive saturations corresponding to this more elaborate pattern, the natural procedure is to undertake a fresh group factor analysis with subdivided group factors based on the new bipolar-matrix. This means seeking a group factor-matrix which will contain seven factors in all instead of the five assumed by the former procedure. The results of a direct analysis on this basis are set out in Table III B.[1]

The figures show quite clearly that the defects in the results obtained with the previous group factor methods arose from the fact that they made no allowance for the presence of the two broad group factors over and above the narrower factors peculiar to the four narrower ' levels.'

[1] The need for this type of analysis was not at first anticipated, although the results of the subdivided bipolar analysis which I carried out at an early stage rather suggested the possibility of some such classification. However, with only two broad group factors, each subdividing into only two smaller group factors, the older formulæ are apt to give indeterminate results, and the newer procedure appears to require some experience for its proper use. I am indebted to Professor Burt for advice at this stage. He suggested in correspondence that the method described in his later article on ' Group Factor Analysis ' (this *Journal*, III, Section C, ' Two Group Factors ') might perhaps be fruitfully employed. I have also to acknowledge the help of Dr. Charlotte Banks, who has given the most generous assistance, not only here, but at every stage of my work.

After seven factors had been extracted in this way there were still a few suggestive residuals (none very large) pointing to the possible influence of a few extremely narrow group factors, due either to a common content in certain tests (as in the verbal, numerical, and other non-verbal types) or to a common test procedure (as in the speed tests). Two small subsidiary factors (VIII and IX) were therefore tentatively added to the factor-matrix.

3. *Subdivided Factors obtained by Rotation.* As before, to allow for the possibility of further and more irregular overlapping, it seemed essential to seek a *complete* group factor-matrix by rotating the original bipolar-matrix in such a way as to obtain a pattern based on that reached by the direct method.

An orthogonal rotation matrix was constructed from the new matrix of subdivided factors. Since the group factors just obtained appeared to be related to the fourth as well as the third summational factor, it was decided to include this fourth factor in the bipolar-matrix to be rotated. On the other hand, the fifth factor seemed largely to be a factor of correlated error, and was therefore omitted. It thus became necessary to construct a semi-orthogonal rotation matrix having nine columns but only four rows. The computation was carried out in the usual way ; and the results of post-multiplying the first four columns of Table II A by this transformation matrix are shown in Table III C. The significant saturations are printed in heavy type. With this final procedure, it will be seen, no significant negative saturations appear : apart from one or two trifling exceptions, all the negative figures are now numerically less than 0·08.

Table III C gives the final answer to our main problem in a fairly clear-cut form. The interpretation of the first seven factors seems plain. As before, the first is a general or basic factor. It shows the highest saturation for the Otis test of intelligence ; and may therefore be identified with the familiar factor of *General Intelligence.*

The second is a broad group factor covering both the two higher levels. It may be termed an ideational or *Intellectual* factor. The tests that it groups together are subdivided by two narrower group factors—the first covering all those belonging to the *Relational* level and the second covering all those belonging to the reproductive or *Associative* level.

There is a second broad factor covering the two lower levels. This may be termed a perceptual or *Practical* factor. The tests grouped together by this factor are also subdivided by two narrower group factors—one for what have been called the *Perceptual* tests (in the narrower sense of the term) and the other for the tests of the simple *Sensori-motor* processes.

The two remaining factors (VIII and IX) have saturations that are barely significant. Their interpretation is decidedly doubtful. Their function, it would seem (at least in part), is to account for the few residual cross-correlations left after the seven main factors have been deducted ; but, since the errors in any set of observed correlations are themselves apt to be correlated, there can be little question that the factors are largely contaminated with these correlated errors. Since such errors are negative as well as positive, their influence would account for the appearance of negative factor-saturations which counterbalance the larger positive saturations that seem to represent distinct and concrete abilities. So far as any interpretation is possible, the positive saturations for the first of these supplementary factors suggest the presence of a verbal or educational factor, the largest saturations being allotted to the tests that are specifically verbal ; the second may conceivably be a factor for tests with a non-verbal content, i.e., those turning on visual imagery and on the manipulation of numbers.

4. *Triangular Factor Analysis.* A hypothesis favoured by early psychologists suggests what might be called a triangular rather than a hierarchical structure. They argued that, since " there is nothing in the intellect that has not previously entered through the senses," the most general capacity of all would be sensory discrimination ; general perceptual ability would be a less general ability entering into all cognitive

processes except sensation ; associative ability would be restricted to the higher levels ; and so on. A hypothesis of this sort was apparently at one time entertained by Spearman. As Burt pointed out, it can readily be tested by the so-called ' triangular group factor method.'[1] The results obtained are plausible enough to indicate that the hypothesis was not so ill judged as its critics declared. But the fit provided is by no means so close as that obtained with the preceding methods.

Comparison of Results. Before deciding which of the various procedures used yields the best result, it seemed proper first of all to examine the average size of the residuals left unfactorized in each case, or, better still, their squares. The poorest fit is furnished by the triangular method ; that obtained with the bifactor-matrix is decidedly better, but still far from satisfactory ; that obtained with the non-overlapping group factors is very much the same. The method of subdivided factors (with one general, one common bipolar, and one subdivided bipolar) yields, with the same number of saturations, a much closer fit than that which is given by the ordinary method of simple summation (with one general and two common bipolar factors). When the method of simple summation is extended to include five factors, a better fit is necessarily obtained than with only four, though the additional amount of variance accounted for is here extremely small. Since the overlapping group factors, obtained by an arithmetical rotation, and the simple structure, obtained by a graphical rotation, were both based on five factors, and since the rotations are orthogonal, the resulting figures of necessity give exactly the same fit as is given by the five summational factors. The subdivided group factors obtained by arithmetical rotation were based on only four summational factors, and hence yield a fit which is not quite so close.[2]

V. GENERAL INFERENCES

Let us now enquire what final inferences may be drawn from the outcome of all these analyses, and what further problems have come to light, calling for more intensive research.

The Most Appropriate Method. With this particular set of tests there can be little doubt that, on both statistical and psychological grounds, the *method of subdivided factors*, supplemented by an arithmetical rotation, provides the most satisfactory means of answering the questions which we set out to solve. First of all, judged from a purely statistical standpoint, regardless of any meaning the factors may have,

[1] This is in effect the traditional method for reducing a square symmetric matrix to a series of orthogonal components. Its introduction into statistics appears to be due first of all to Udny Yule ; but several of Pearson's followers, notably H. B. Heywood (' On Finite Sequences of Real Numbers,' *Proc. Roy. Soc.*, A, CXXXIV, 1931, pp. 486–501), as well as Burt, have discussed its use in factorial work. The working procedure adopted was that described in this *Journal*, IV, 1951, pp. 40–4. With the present correlation table it appears to yield no new information ; and the results are of academic interest only. They illustrate in the concrete one early conception of hierarchical structure, namely, that in which the mental processes of the *simplest* level are identified with the general factor, as suggested, for example, by Spearman in his first paper on the subject. For a further discussion of this particular version of the theory I may again refer to my thesis (18).

[2] Some writers appear to suppose that the choice of a group factor-matrix rather than a factor-matrix involving either bipolar factors or (it may be) simple structure is to be decided *solely* from an examination of the closeness of fit obtained with the alternative procedures. That of course is scarcely correct. If the orthogonal rotation matrices yielding the group factors and the simple structure are both of the same order and are both applied to the same summational or centroid matrix, all three procedures must necessarily lead to a reconstructed correlation table which is precisely the same and therefore yields precisely the same fit for the observed correlation table. However, with *unrotated* factors the closeness of fit obtained with a given number of significant saturations does afford one important method of evaluating different factor-patterns. But the factor-pattern itself ought in general (like other forms of distribution) to be postulated as part of the prior hypothesis to be tested ; the statistical analysis can do little more than state that the posterior data are, or are not, consistent with that postulate.

it furnishes as close a fit as can be reasonably demanded with a sample of this size, and at the same time it reduces the negative saturations to non-significance without either substituting correlated for uncorrelated factors or introducing uncorrelated factors with highly correlated saturations. But secondly, the procedure enables us, in the most direct and objective fashion possible, to decide between the rival schemes set forth in our initial hypotheses. The pattern of figures to which the method leads is based solely on mathematical considerations. Thus, without in any way begging the question or using the scheme to be verified as a guide to the mode of rotation, we arrive at a detailed classification of the mental processes tested.

The Most Appropriate Hypothesis. As it turns out, the classification thus automatically reached agrees in all essential points with that put forward in the scheme we undertook to test.

To begin with, what we discover is, not a miscellaneous collection of sporadic ' abilities,' but a systematic series of cognitive capacities for tasks of increasing complexity, developing apparently by stages, such as we might anticipate from evolutionary and developmental considerations. In the broadest possible sense, therefore, the scheme of factors found is certainly ' hierarchical.'

Secondly, when we look more closely into the nature of the scheme, we find, first, that there is a single basic factor common to all cognitive processes ; next, that these processes are divisible into two broad factors—the intellectual and the practical, as we have provisionally termed them ; and then each of these broad factors is itself subdivided into two narrower factors, which go far to explain the traditional notion of the four cognitive levels ; finally, these narrower factors in their turn tend to split into factors still narrower, varying according to the form or the content of the tests and spreading in some cases over two or more levels.

The alternative conceptions of hierarchical structure, which we found were still adopted by various writers, appear to be decisively disproved.

(*a*) The ' unifocal ' view, so attractive by its very simplicity, which is usually attributed to Spearman and which Spearman himself attributes to Spencer, is assuredly refuted beyond any shadow of doubt : in the light of our results it is quite impossible to maintain that cognitive abilities are manifestations of a single general factor only, showing a simple " serial arrangement in proportion as intelligence advances." Hence we can confidently endorse McDougall's verdict on Spearman's two-factor theory : " such a view is much too simple to interpret the facts disclosed by the correlations." [1]

(*b*) The ' multifocal ' view, formulated most clearly by Thorndike, and retained by many of his followers after Thorndike himself had apparently abandoned it, is also unsupported by the results. In its most definite form this assumed that cognitive abilities were of three distinct types, belonging to different and successively emerging stages of development, without any general factor common to them all. Thurstone's ' simple structure ' is a more recent version of the same general type : this seems equally untenable.

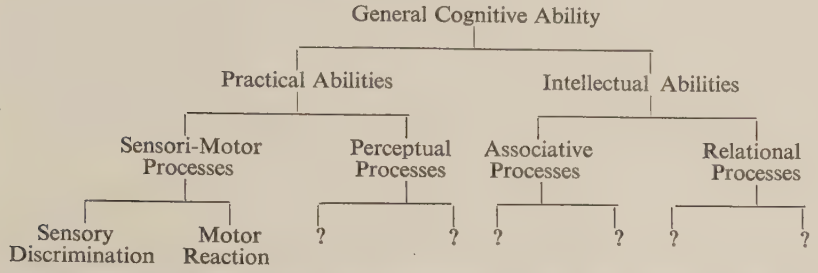
Having eliminated all the main alternatives we are consequently left with the theory that mental activities reveal a hierarchical structure in the full and technical sense of that term. To quote McDougall once again : such a ' basis ' for the ' mind's structure,' though founded originally on the genetic study of human and animal life, and reinforced by later neurological work, appears to be " in agreement with, and so supported by, the analysis of mental functions as achieved by mental testing and the statistical examination of data from a wide variety of tests." [2] The main

[1] *Energies of Men*, 1932, p. 94. McDougall goes on to suggest an interesting correlational research which would, he believes, corroborate his view : it is curious that no one has hitherto thought of carrying it out.

[2] *Loc. cit.*, p. 91 : cf. also p. 95. McDougall adds that " Professor Spearman himself, as well as his critics, seems to be moving in the direction of such a view." The genetic argument is confirmed by applying the tests to younger children, of different ages but similar I.Q. The tests of the lowest level are readily performed by the youngest, but those of the highest level cannot be performed at all. However, to demonstrate the parallel in detail would require a separate research.

features of the final scheme may be most readily grasped from the genealogical tree or stemma set out in Table IV.

TABLE IV. HIERARCHICAL SCHEME OF FACTORS



Note. The queries in the bottom line are intended to suggest the probable existence of dichotomous divisions into afferent and efferent factors, with further subdivisions (or cross-divisions) into factors of content.

The General Factor. The chief features of the scheme call for a word or two of comment, and raise further problems of their own. The result which is perhaps most striking, since it emerges with each of the alternative methods of analysis, is the confirmation of a general factor underlying all the processes tested. Every effort to suppress it, with a view to securing ' simple structure,' failed. Since the tests were deliberately selected to represent all the more important forms of cognition, we may infer that this factor represents a general cognitive capacity, identifiable with what Spencer called ' intelligence ' and Galton ' general ability '—an identification confirmed by the external criterion furnished by the Otis test.

Strictly speaking, to prove that this comprehensive factor is not susceptible of a still wider designation, we ought to have included in our battery measurements of non-cognitive as well as cognitive aspects, and then shown that the significant saturations were confined solely to the latter. This seemed scarcely necessary in a preliminary enquiry ; but it is a point that deserves further attention in future research.[1]

The relative size of the different saturations for the general factor itself should be noted. With the methods here used, what are called ' saturation coefficients,' it should be remembered, specify the correlations between the several tests and the factor in question. As will be seen from the tables, apart from a few minor exceptions, the size of the coefficients is alone almost sufficient to indicate to which level each test belongs. With some of the analyses, it is true, the levels appear to overlap a little ; but with others (e.g., the first) the distinction into levels is clearly marked by definite gaps in the sequence of figures. Such gaps, however, may be partly due to the method adopted in selecting tests at the very outset ; special care was taken to avoid any form of test that might involve processes from two different levels to much the same degree, and so fall between adjacent classes.[2]

[1] This gap in the logic of the usual proof has been pointed out by Burt. In one or two researches he himself has included assessments of affective and conative characteristics. He finds that, although the so-called general cognitive factor often showed positive saturations for these other aspects, they were extremely small and often non-significant. These and other results have led him to question whether the antithesis ' cognitive *versus* orectic ' is the best way of distinguishing a factor which includes ' practical ' efficiency as well as purely cognitive efficiency in the narrower sense : (cf. ' The Assessment of Personality,' *Egypt. J. Psychol.*, II, 1946, pp. 4f., and refs.). It would certainly appear that there may be a factor for mental efficiency of still wider generality than the so-called general cognitive factor—a factor (probably quite small) which would include efficiency of character as well as efficiency of intelligence ; and this perhaps may account for Spearman's preference for identifying his *g* with general mental energy and Garnett's suggestion that it might be identified with will.

[2] A critic, who has kindly read my arguments in manuscript, objects that, " in view of this deliberate selection of tests, all that is shown by the factor analysis is that the selection was successful : to prove your point " (he adds) " should you not have applied an unselected set or else a sample selected at random ? As it is, what you have verified are clear-cut levels in your battery of tests, not in the structure of the mind." In reply let me first point out that what were selected were not tests but mental processes ; and in any case, unless the suspected differences actually exist in the structure

But, whether or not the levels do in fact overlap or even merge, one conclusion stands out quite clearly : " the more complex the process tested, the closer the correspondence with intelligence." [1] This is what we might anticipate as a consequence of the hierarchical theory as ordinarily tested. Spencer, still thinking in terms of the old associationist school, put forward a definite criterion of level : the higher and more developed processes are " those which show the more *complex correspondences* between the environmental situations that evoke them, and the mental or motor reactions to which they lead." [2] Much the same notion was adopted by Sherrington in his general picture of the working of the central nervous system, and expressed in more technical language. " The central nervous system," he writes, " is the meeting place where afferent paths conjoin and conduct to an efferent path : in short, an organ of co-ordination. The mechanisms for the common (outgoing) path knit together, by pluri-receptive summation, separate stimuli of similar or of different kinds. The *degree to which paths are confluent* is thus the touchstone for the rank of a centre in this neural hierarchy. Hughlings Jackson's ' three levels ' is an expressive figure for this grading of rank " ((2), pp. 313–14). All this bears out Burt's interpretation rather than Spearman's.

One or two suggestive corollaries appear to follow. If the complexity of a cognitive process is assumed to depend upon the complexity of the neural processes involved, and if the complexity of the neural processes (i.e., the number of confluent paths giving rise to them) depends on the level of the organizing or co-ordinating centre within the central nervous system, then we might infer that the saturation for the general factor would provide an index of the level of that process in the psychoneural hierarchy—i.e., " its height in the scale of mental development," and so determine the " weight " which should be assigned to it in assessing general psychoneural efficiency ((3), p. 104).

But presumably the degree of complexity indicated by the saturation for the first factor is (as Burt remarks) merely *general* cognitive complexity ; to assess the *total* complexity of the process we ought strictly to take into account the contributions of the smaller factors as well. Total complexity might therefore be measured by the self-correlation, i.e., the sum of the squares of the saturations for all the verifiable factors. This suggests further possibilities. One of the most urgent needs of individual psychology is an objective unit of measurement. At present the unit most widely used is the standard deviation, i.e., the square root of the variance. Now the common factor variance of any test is generally equated with its (reduced) self-correlation ; and, if this depends on the total complexity of the psychoneural processes involved, some convenient and perhaps conventional measure of complexity might be derived from it which would furnish an objective mental unit. Here is yet another problem awaiting research.[3]

It further follows that ' intelligence ' is not to be conceived as an unanalysable ' faculty '—a ' primary ability ' or ' unitary trait.' It is a formal rather than a material capacity. It has no special organ or centre, but simply represents the degree to which the individual's mental life, at least on its cognitive or rather its directive side, is organized to make adjustments that are at once intricate and unified.

of the mind, it would be impossible to demonstrate them either in the tests or in the test-results. Suppose that the hypothesis to be verified had been the presence of distinct geological layers in the Derbyshire Peak. Merely to pick up a score of fragments just at random would not be a very satisfactory procedure ; we might easily miss important types. If, however, casual observation had already suggested the presence of four distinct minerals—sandstone, limestone, carbon, and chalk—the first step would obviously be to find half-a-dozen specimens of each suspected type ; and then (assuming the naked eye was not to be trusted) devise appropriate chemical tests for each—a reagent for silicon, another for calcium carbonate, another for carbon, and so on. To select twenty tests at random might easily be fruitless. Nor will the fact that we choose our various tests with a definite hypothesis in view produce differences of structure in the mountain if those differences are not already there. I agree that, in following up this preliminary research, a wider variety of mental tests will be desirable to see whether there may not be further differences hitherto unsuspected and to determine how sharp the differences may be separating one type of process from the next. Further, an entirely new method of approach would be needed—applying the tests to children at successive ages—if we are to verify the order of the several levels.

[1] Burt (3), p. 169, (6), p. 95. This was one of the chief reasons why Burt and others have maintained that intelligence represents, not simply a discriminative or analytic ability (as Spearman originally argued), but also an integrative or synthetic process. In defending his view, Spearman answered that such complexity did not imply any ' synthetic power,' but merely allowed " the multitude of independent specific factors to cancel one another, so that the general factor (mental energy) remains more dominant " ((7), p. 69).

[2] See *Principles of Psychology*, I, p. 350.

[3] There appears to be some experimental evidence for this view. As Burt observes, with tests of ' reaction-time ' involving reactions of increasing complexity, both the standard deviation and the correlation with the general factor seemed roughly proportional to the complexity of the process : cf. *Factors of the Mind*, pp. 283–4 and refs., for the evidence and for a further discussion of the metric possibilities of this view. The same idea seems to underlie Thomson's ' sampling theory ' : " the communalities " (he says) " indicate the fraction which each test forms of the whole mind."

Nor does it depend on the mere ' quantity of connexions ' or the mere number of ' bonds ' or ' ideas ' within the individual's mind, which an intelligence test will proceed blindly to ' sample,' [1] it depends on an organized selection, rather than on a random selection, of ' connexions,' ' bonds,' or ' ideas.'

The Subdivided Group Factors. But, while recognizing the predominance of the ' general factor,' in the scheme set out above, there can be no question about the existence and importance of the supplementary factors. Of these the most conspicuous (possibly because the experiment was planned with this feature primarily in mind) appear to be the factors for the four main levels. It is instructive to observe that the distinctions between these levels come most clearly to light, not in the commoner types of rotated factor-matrix, but in the various forms of bipolar analysis. We must not, however, take the four-level scheme as a fixed and final picture of the essential organization of psychoneural structure. The wider classification indicated by the two broader factors is a sufficient reminder of this point. Nor indeed can we expect that a growing dynamic system will conform to the simple static outlines that a human architect might have laid down. The most that the factorist seeks to achieve is merely a schematized model, approximating as closely to the relevant facts as his limited data will permit. And the adoption of a theory of *subdivided* factors will enable him to re-classify and to sub-classify almost as much as his particular purpose may require or future data may justify.

In any further investigation it will plainly be desirable to examine more precisely the relations between the familiar ' material ' factors (those turning on various *contents*—verbal, numerical, visual, auditory, kinæsthetic, and the like) and the more ' formal ' factors (those that depend on the general nature of certain mental *processes*, sensory discrimination, perception, association, and the rest). In particular the proper interpretation of the two broader factors just referred to calls urgently for further research. In borrowing the popular terms ' intellectual ' and ' practical,' it was apparently assumed that the distinction is, as it were, vertical rather than horizontal, the consequence of the organization of the psychoneural scheme into afferent and efferent aspects rather than into ' perceptual ' and ' conceptual ' layers or levels. However, there is no necessary incompatibility between the two metaphorical modes of description. Purely perceptual activities tend to issue almost immediately in efferent or motor action, and hence allow little time for processes of memory and the reproduction of images and ideas, or for a reflective analysis in conceptual and relational terms. On the other hand, an appeal to these latter activities tends inevitably to inhibit an immediate practical response.

The Relation between Cognition and Orexis. The restriction of such investigations as the present to the cognitive or directive side is itself somewhat arbitrary, arising from the investigator's need to confine the problem of the moment within practicable limits. The distinction between practical (or perceptual) and intellectual (or reflective) activities is suggestive of an analogous distinction appearing on the orectic or energetic side, namely, the contrast between what Binet termed ' objective ' and ' subjective ' attitudes and what Jung has renamed the ' extravert ' and ' introvert ' reactions.[2] A wider enquiry, therefore, covering non-cognitive as well as cognitive characteristics, seems necessary before we can fully understand the nature of the broader factors here revealed. Moreover, far more factorial work is needed on non-cognitive qualities in and for themselves before we can confidently

[1] Thorndike, E. L. (9), pp. 415–18 ; Thomson, G. H., *The Factorial Analysis of Human Ability*, pp. 45, 47, 54, and 304f. Thorndike's ' quantity hypothesis ' is expressed in terms of the number of " connexions, associations, or bonds which can be formed between ideas," and an ' idea ' is defined as " any small portion of the stream of thought, such as a percept, image, meaning, or word." Thomson similarly speaks in terms of ' bonds.' In his section on ' the inequality of men.' he says each man may be regarded as a sample of all the possible bonds : " each possesses some, but not all, of the neural bonds which are the physical side of thought ; some are rich, others poor, in these bonds." " Each test calls upon a sample of the bonds." Hence in his view the appearance of a hierarchical arrangement has " nothing to do with the structure of the mind." The mind itself is " a comparatively undifferentiated complex " ; and the analysis into factors is " a result of the laws of chance, not of any psychological laws." However, a theory of cognitive ability which apparently dispenses with both differentiation and integration (or at least treats those characteristics as relatively unimportant) would seem to be, not ' hierarchic,' but (in Spearman's phrase) ' anarchic.' I do not wish to deny that chance may play some part, and often a large part, in determining the amount of correlation between pairs of tests, but only that it plays the main part.

[2] Cf. (15), pp. 109–10, and *The Measurement of Mental Capacities*, 1927, p. 36.

accept the view that a hierarchical type of organization characterizes the whole range of mental life.[1]

All these are issues which can hardly be solved by statistical analysis alone. As we saw at the outset, the more plausible hypotheses, between which we endeavour to decide, must be carefully formulated in the light of the latest non-statistical, as well as statistical, investigations. Probably any final verification will be most easily secured by an application of statistical analysis to the results of planned experiments designed from the outset to permit a statistical verdict. The logical superiority of this more eclectic mode of approach has, it is hoped, been plainly demonstrated in the brief and tentative enquiry reported in the preceding pages.

The General Value of the Conception. The view that the potentialities of the mind (or perhaps we should rather say of the underlying central nervous structures) are organized in a hierarchical system is in itself an exceedingly fruitful conception. To begin with, it explains many outstanding facts that have come to light since the idea was first clearly formulated and applied in factorial work. Nothing is more instructive than to select a few of the earlier correlation tables put forward in support of some one or other of the older factor theories, and re-analyse them by the method of subdivided factors : it is surprising to find how much more completely their apparent peculiarities are then explained than if we fall back on the two-factor theory of Spearman, the sampling theory of Thomson, or the theory of simple structure proposed by Thurstone. The trouble with most of these theories is that they were put forward solely or mainly with statistical considerations in view. Had the mind been constructed on the extremely simple lines assumed by the two-factor theory or in the haphazard undifferentiated way required by the sampling theory, or had it consisted merely of a miscellaneous and unsystematized collection of ' primary abilities,' it is difficult to see, from a biological standpoint, how individuals and species so equipped could have either survived or evolved.

The conceptual scheme here described has much wider implications. It not only provides a convenient basis for work in individual psychology : it also illuminates much that is otherwise obscure or unexplained in the field of general psychology. It helps to co-ordinate numerous isolated observations on mental growth and adult behaviour, and elucidates many of the experimental results obtained in studies of maturation, learning, and thought processes. Above all, as a conceptual scheme, it immediately introduces something like an intelligible order into the chaotic list of topics that figure in current psychological text-books ; at the same time it suggests numerous deducible corollaries of interest in themselves and capable of direct statistical checks by further research.

VI. SUMMARY AND CONCLUSIONS

1. In its more general form the view that ' the mind has a hierarchical structure ' was originally put forward on the basis of introspective, evolutionary, and neurological arguments, and, according to Burt and others, has been, or at any rate might be, confirmed by statistical analysis. Many recent writers, however, have rejected the notion of anything approaching a systematic structure, and have maintained that the mind consists either of a miscellaneous set of ' unitary abilities ' or else of an ' undifferentiated complex.' Those who accept the idea of hierarchical organization interpret it in different ways : it is taken to imply either (i) a *divergent* classification, involving a general factor subdivided into group factors of progressively diminishing

[1] For this purpose, it would seem, the method of subdivided factors will require an additional complication. According to McDougall's diagram (*Social Psychology*, p. 440) the ' self-regarding factor ' is a general factor extending over all moral activities ; the other sentiments are limited to certain groups ; and the primary emotions may be represented by still narrower group factors. But each of the latter is capable of entering into many if not all of the former. For a further discussion of the problem, see ' The Factorial Study of the Emotions,' *ap. Feelings and Emotions, loc. cit. sup.*, pp. 531f.

range (McDougall, Burt), (ii) a serial or *unifocal* classification, involving one general factor only (Spearman), or (iii) a *multifocal* classification, involving group factors for separate levels, with no general factor (Thorndike, Thomson). The object of the present investigation was to determine whether, by means of an appropriately planned experiment and an appropriate type of statistical analysis, a more decisive verification of one or other of these alternative hypotheses could be secured.

2. A set of 20 tests was employed, each constructed so that its results would depend, as far as possible, on the activities of a single cognitive level only. They were then applied to 166 boys, aged 10.0 to 11.0 years. The correlation table was analysed by all the main methods of factor analysis, namely, in terms of bipolar factors, of simple group factors, and of rotated group factors, both ' undivided ' and ' subdivided.' Thurstone's method of ' simple structure,' Holzinger's ' bifactor method,' and Yule's ' triangular method ' were also tried.

3. A comparison of the various procedures indicates that the method of *subdivided group factors*, obtained by an arithmetical rotation from a matrix of subdivided bipolar factors, is by far the most appropriate for such researches as the present. All the methods revealed a large general factor roughly identifiable with general cognitive ability or *intelligence*. Both the bipolar and the group factor analysis indicated the presence of two broad group factors provisionally identified with (i) *practical* and (ii) *intellectual* abilities respectively ; these were found to subdivide into narrower group factors for (i*a*) *sensori-motor*, (i*b*) *perceptual*, (ii*a*) *associative*, and (ii*b*) *relational* processes respectively. There were some indications that these narrower factors were still further subdivisible according to either the *content* or the *form* of the test. The most satisfactory fit to the data was obtained with a matrix of ' subdivided ' group factors corresponding to the broad, narrower, and narrowest groups indicated by the bipolar analysis.

4. It was concluded that Burt's hypothesis of a hierarchical structure, involving a divergent classification of mental characteristics into subdivisible groups, is fully confirmed by the statistical evidence, and may be further supported by genetic and neurological evidence. It is, however, emphasized that the factorial scheme thus suggested should be regarded as no more than an approximate theoretical model and as a suggestive foundation for future investigations of a more extensive and a more intensive type. Such investigations should be based on a *combination* of factorial techniques with other modes of approach, not on statistical analysis alone. If finally confirmed, the conception is likely to prove of value, not merely to those who specialize in mental tests or factorial theories, but for the whole field of general and individual psychology.

REFERENCES

1. McDougall, W. (1905). *Physiological Psychology*. London : Dent.
2. Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. London : Constable.
3. Burt, C. (1909). ' Experimental Tests of General Intelligence.' *Brit. J. Psych.*, III, pp. 94–177.
4. Burt, C. (1910). ' General and Special Abilities in School Children.' *Brit. Ass. Ann. Rep.*, 1910, pp. 804–5.
5. Thorndike, E. L. (1910). *Educational Psychology*. New York : Columbia University.
6. Burt, C. (1911). ' Experimental Tests of Higher Mental Processes and their Relation to General Intelligence.' *J. Exp. Ped.*, I, pp. 93–112.
7. Hart, B., and Spearman, C. (1912). ' General Ability : its Existence and Nature.' *Brit. J. Psych.*, V, pp. 51–84.
8. Thorndike, E. L. (1923). *Educational Psychology*, Vol. III : *Individual Differences and their Causes*. New York : Columbia University.
9. Thorndike, E. L., *et al.* (1926). *The Measurement of Intelligence*. New York : Columbia University.
10. Burt, C. (1927). *The Measurement of Mental Capacities*. Edinburgh : Oliver & Boyd.

11. McDougall, W. (1932). *The Energies of Men.* London : Methuen.
12. Cattell, R. B. (1946). *The Description and Measurement of Personality.* New York : World Book Company.
13. Eysenck, H. J. (1947). *The Dimensions of Personality.* London : Kegan Paul, Trench, Trubner.
14. Wilkie, J. S. (1948). ' The Biological Approach to the Mind-Body Problem.' (Unpublished Ph.D. Thesis, University of London Library.)
15. Burt, C. (1949). ' Subdivided Factors.' This *Journal,* II, pp. 41–63.
16. Burt, C. (1949). ' The Structure of the Mind : a Review of the Results of Factor Analysis.' *Brit. J. Educ. Psychol.,* XIX, pp. 100–14, 176–99.
17. Vernon, P. E. (1950). *The Structure of Human Abilities.* London : Methuen.
18. Moursy, E. M. (1952). *A Factorial Investigation of the Hierarchical Organization of Cognitive Levels.* (Unpublished Ph.D. Thesis, University of London Library.)

Vol. V
Part III

The British Journal of Psychology,
Statistical Section

November,
1952

AN ITERATIVE METHOD FOR COMPUTING INVERSE MATRICES

By MAX HAMILTON

Department of Psychology, University College, London

I. *Introduction.* II. *Iterative Methods.* III. *The Present Method.* IV. *Worked Example.* V. *Summary.*

I. INTRODUCTION

This paper describes a new method for calculating inverse matrices, particularly those arising in statistical work, e.g., dispersion and correlation matrices.[1] It shows how an initial matrix may speedily be obtained for an iterative procedure of the usual type, and how the convergence of the iterations during the initial stages may be accelerated to produce a satisfactory inverse with little work.

There has of late been a renewed interest in computational methods. Two factors have played an important part, the development of multivariate analysis and the improvement in calculating machines. High-speed electronic calculators have achieved great prominence in the press. But for most of the computational work in current research, they might as well not exist. In their present form they are scarcely suitable for the psychologist's needs ; and their use is still largely restricted to secret work. Desk calculators have undergone great improvements : they have become faster, less liable to break down, and now include devices for obtaining sums of products and quotients, for continued multiplication, and for squaring. Thus most of the computational labour involved in multivariate analysis has been made mechanical and speeded up. As a result it is now worth while devising fresh methods of computing which are simpler to apply than the old, even though they may demand more numerous elementary operations, like addition, multiplication, etc., such as can to-day be carried out by a machine.

II. ITERATIVE METHODS

Of these simpler procedures, those based on iteration are among the most useful. No doubt they suffer from certain disadvantages : occasionally it is difficult to obtain results of requisite exactitude, and so in the end the repetitions may take longer than a direct procedure. Yet they have obvious merits. First, they satisfy the natural desire to reach a preliminary notion of the answer at an early stage. Secondly, the progress of the iteration automatically reveals and rectifies mistakes. Thirdly, the extra time required for the iterations is rapidly diminished as improved mechanical computers become available. In practice direct methods seldom possess the advantages that a superficial impression may suggest ; and, whereas iterative methods start by working with only a very few significant figures (often only one or two), the direct

[1] When I submitted the first draft of this manuscript to Professor Burt, he showed me a manuscript containing six methods for calculating inverses for correlation matrices, including certain suggestions which come near to the procedure described below. Had it been possible, I should have preferred both to have been published simultaneously ; and in any case hope that Professor Burt may find time to complete his for the press in the near future. The equations with which the following proof starts (eqn. (1) and (3), p. 182 below) are given in his recent paper in this *Journal* ((5), Sect. V, ' Inverses '), where the matrix procedure is deduced by analogy with a familiar device used in elementary algebra for obtaining reciprocals to scalar numbers by an iterative procedure.

method requires a large number from the very outset. It is true that, if with the direct method we begin with a smaller number, it is nearly always possible to determine the size of the errors produced by rounding ; but then, if it is found that they are excessive, either the whole calculation has to be repeated with a larger number of figures, or else the result has to be improved by iteration. This, as Hotelling points out, is one of the most powerful arguments for iterative procedures.

These points apply with particular force to the inversion of matrices. As the size of the matrix is enlarged, the number of operations and the errors of rounding increase out of proportion. Several direct methods have been developed during recent years, but all are laborious. For cases in which an approximate inverse is already available, Hotelling (1) gives details for determining its accuracy, and the number of iterations required to reach any desired degree of accuracy. Unfortunately it is necessary to start with an approximation which is not too remote ; otherwise convergence will be extremely slow or non-existent.

Hotelling himself has proposed two methods for dealing with such cases. In his 1943 paper he suggests that a direct method be used first, taken to two figures only. In a later paper (2) he discusses the possibility of preparing tables of inverses for matrices of standard types. There are at least two methods for deriving the inverses of matrices of high order from those of a lower ; but with the limitations proposed, a 3×3 matrix of single digits would require 715 inverses. Thus even the initial stages would entail considerable work.

This paper shows how a first approximation can be found by inspection, although in practice the computer may usually begin with the second iteration. A method is then given for accelerating the convergence ; and this is effective however remote the first approximation. Finally, an example is given of a device for dealing with laggard elements.

III. THE PRESENT METHOD

The following theorem has long been known, and the standard proof will be found in such text-books as (3). The proof given below is based on the analogy between scalar and matrix operations pointed out by Burt (5). If A be the given matrix, A^{-1} its inverse, C_0 its approximate inverse, and I the unit matrix, then we define

$$D_0 = I - AC_0 \text{ and } C_1 = C_0(I + D_0).$$

It follows at once that $$C_1 = C_0(2I - AC_0). \qquad (1)$$

Now from the definition $AC_0 = I - D_0$;

therefore $$A^{-1} = C_0(I - D_0)^{-1}. \qquad (2)$$

We may accordingly regard $(I - D_0)^{-1}$ as a correcting factor which converts C_0 into A^{-1}.

But $$(I - D_0)^{-1} = I + D_0 + D_0^2 + \ldots + D_0^{n-1} + D_0^n (I - D_0)^{-1}. \qquad (3)$$

Here the right-hand side forms a converging series, provided the latent roots of D_0 have a modulus less than 1. In that case, $(I + D_0)$ may be used as a first approximation to $(I - D_0)^{-1}$. Hence, by (1), C_1 is an approximation closer to A^{-1} than C_0. A repetition of the iterative process represented by (2) is equivalent to taking more of the terms from the right-hand side of (3).

Now, simply to apply this process blindly and automatically over and over again will no doubt lead eventually to the correct solution. But the calculations will be long and tedious. The special feature of the working method I propose arises from the exploitation of certain incidental peculiarities to accelerate the approach to the final result.

Consider the typical dispersion matrix : its elements are variances and covariances, the latter being in general smaller than the former. And to begin with let us confine ourselves to the case where all the values are positive. If the variables were uncorrelated, the covariances would be zero and the matrix a simple diagonal matrix. Its inverse is found by changing

the elements of the diagonal to their reciprocals. If the covariances be not zero but quite small, this inverse will still remain as an approximation to the exact inverse.

Let the dispersion matrix be denoted by A, and the elements by a_{ij} ($i, j = 1$ to n). If it be post-multiplied by the approximate inverse, the product will be

$$AC_0 = \begin{bmatrix} 1 & a_{12}/a_{22} & \dots & a_{1n}/a_{nn} \\ a_{21}/a_{11} & 1 & \dots & a_{2n}/a_{nn} \\ \dots & \dots & \dots & \dots \\ a_{n1}/a_{11} & a_{n2}/a_{22} & \dots & 1 \end{bmatrix}, \qquad (4)$$

the elements being unity in the leading diagonal and a_{ij}/a_{jj} (a fraction less than unity) elsewhere. The original guess can now be improved by iteration. The above product matrix is first subtracted from $2I$ (which leaves the leading diagonal unchanged, but changes the sign of the other elements), and the remainder then pre-multiplied by the original guessed inverse. The new approximate inverse is

$$C_1 = C_0(2I - AC_0) = \begin{bmatrix} 1/a_{11} & -a_{12}/a_{11}a_{22} & \dots & -a_{1n}/a_{11}a_{nn} \\ -a_{21}/a_{11}a_{22} & 1/a_{22} & \dots & -a_{2n}/a_{22}a_{nn} \\ \dots & \dots & \dots & \dots \\ -a_{n1}/a_{11}a_{nn} & -a_{n2}/a_{22}a_{nn} & \dots & 1/a_{nn} \end{bmatrix}. \qquad (5)$$

If the dispersion matrix is a correlation matrix, with unity in the leading diagonal, then this new approximation, C_1, is the same as the original correlation matrix with the signs of the coefficients reversed; it can thus be written down at once by inspection.

If we now form the product AC_1, this will provide a still closer approximation to I than AC_0.

$$AC_1 = \begin{bmatrix} 1 - \frac{a_{12}\cdot a_{21}}{a_{11}\cdot a_{22}} - \dots - \frac{a_{1n}\cdot a_{n1}}{a_{11}\cdot a_{nn}}, & -\frac{a_{11}\cdot a_{12}}{a_{11}\cdot a_{22}} + \frac{a_{13}}{a_{22}} - \dots - \frac{a_{1n}\cdot a_{n2}}{a_{22}\cdot a_{nn}}, \dots & -\frac{a_{11}\cdot a_{1n}}{a_{11}\cdot a_{nn}} - \frac{a_{12}\cdot a_{2n}}{a_{22}\cdot a_{nn}} - \dots + \frac{a_{1n}}{a_{nn}} \\ \frac{a_{21}}{a_{11}} - \frac{a_{21}\cdot a_{22}}{a_{11}\cdot a_{22}} - \dots - \frac{a_{n1}\cdot a_{2n}}{a_{11}\cdot a_{nn}}, & -\frac{a_{21}\cdot a_{12}}{a_{11}\cdot a_{22}} + 1 - \dots - \frac{a_{2n}\cdot a_{n2}}{a_{22}\cdot a_{nn}}, \dots & -\frac{a_{21}\cdot a_{1n}}{a_{11}\cdot a_{nn}} - \frac{a_{22}\cdot a_{2n}}{a_{22}\cdot a_{nn}} - \dots + \frac{a_{2n}}{a_{nn}} \\ \dots\dots\dots\dots & \dots\dots\dots\dots\dots & \dots \quad \dots\dots\dots\dots\dots \\ \frac{a_{n1}}{a_{11}} - \frac{a_{21}\cdot a_{n2}}{a_{11}\cdot a_{22}} - \dots - \frac{a_{n1}\cdot a_{nn}}{a_{11}\cdot a_{nn}}, & -\frac{a_{n1}\cdot a_{12}}{a_{11}\cdot a_{22}} + \frac{a_{n2}}{a_{22}} - \dots - \frac{a_{nn}\cdot a_{n2}}{a_{22}\cdot a_{nn}}, \dots & -\frac{a_{n1}\cdot a_{1n}}{a_{11}\cdot a_{nn}} - \frac{a_{n2}\cdot a_{2n}}{a_{22}\cdot a_{nn}} - \dots + 1 \end{bmatrix}$$

Inspection reveals two important characteristics. Consider the first column: its first element contains a 1 from which is subtracted a series of fractions; each of these (assuming positive covariances) will be positive. This first element is, therefore, less than 1, but approaches it when the covariances approach zero. The other elements consist each of n fractions, the first of which is positive, the others being negative. One of them always equals the first; in the second, the second fraction equals the first; in the third the third; and in general the jth fraction is equal to the first fraction in the jth element. The two equal fractions vanish, leaving the $n - 2$ fractions all negative. All the elements of the first column, except the first, are therefore negative; and the first element is always less than 1. The result is similar for the other columns. The jth column has its jth element (the element in the leading diagonal) less than 1, and all its other elements negative.

Returning now to the first column of AC_1: each of its elements is the inner product of the corresponding row of A with the first column of C_1. Let a positive quantity x_1 be added to the first element of this column such that the inner product with the first row of A now equals *one*. The effect is to add to each element in the first column of AC_1 a positive quantity equal to $a_{j1}x_1$ in the jth element: (as a rule, $a_{j1}x_1$ will be less than $a_{11}x_1$). This added number, x_1, will affect only the first column of AC_1.

A similar change is now made in the second column of C_1 by adding to the second element a quantity x_2, of a size such that the second element of the second column of AC_1 is made equal to 1. And so on. In matrix notation we add to the matrix C_1 a diagonal matrix X; and the product, $A(C_1 + X) = AC_1 + AX$, is evidently made up of AC_1 with the matrix AX added to it. Since X is a diagonal matrix, AX is formed by multiplying each of the columns of A by the appropriate element of X. Its elements are positive. Hence, when added to AC_1 it raises the elements in the leading diagonal to unity, and brings the other elements nearer to zero. The resulting matrix is consequently brought nearer to I; and thus C_1 is brought nearer to A^{-1}. The next step follows the same iterative routine. The new matrix, $AC_1 + AX$,

is subtracted from $2I$; and the remainder pre-multiplied by the modified C_1 to produce C_2. The cycle is repeated as often as is necessary ; and a highly accurate inverse is quickly attained.

To make both the derivation and the method clearer two assumptions were made at the start, namely, that the covariances were small compared with the variances, and that they were all positive. If the covariances are large, convergence will be slower. Nevertheless, the method of adding a suitable quantity to the elements of the leading diagonal overcomes this slowness without appreciably increasing the labour. If some of the covariances should be negative, this will tend to make the first form of AC_1 nearer to I. If a large number of the correlations were negative, then the x_j to be added might have to be negative. (It is likely that such a matrix would not have an inverse.) If we assume that the covariances are all negative, then clearly the elements of AC_1 will all be positive, and those in the leading diagonal will be greater than 1. In either extreme case, the structure of AC_1 is such that whatever will bring each element of its leading diagonal towards 1 will bring its other elements towards zero. It may even swing them beyond zero ; but this will not affect progress.

The novel features in the procedure I have described are so flexible as to demand a verbal rather than a formal or symbolic description. Hence a proof of convergence cannot follow the usual formal style. As is well known, an appearance of convergence may be deceptive. The classic example is the series $1/2 + 1/3 + 1/4 + \ldots$, which does not converge despite its appearance. Although the *sum* does not converge to a limit, the individual *elements* do. They converge to zero because we have deliberately made them or chosen them to do so. This argument applies here. In the method proposed, we can examine the matrix AC_1, and then use our information to alter C_1 in the direction we wish it to go. Whatever the difference between C_1 and A^{-1} we deliberately diminish it. Richardson has pointed out (4) that a formal proof of convergence may not be helpful if it implies that the required result will be attained only after an infinity of computations.

In actual practice, the matrix C_1 will be obtained directly from A without any of the intermediate steps ; and its leading diagonal will be left blank. AC_1 need not be explicitly computed. The leading diagonal of AC_1 and of C_1 are obtained simultaneously in the final modified form ; then the modified AC_1 is computed for the iterative cycle. Further, it is convenient to build up the elements of the leading diagonal to a little above *one*. With matrices of moderate size, say, of order 4 to 8, the first two trials need only be taken to one, or at most two, significant figures ; with larger matrices it is well to use two from the start. But the number should not be increased until the results settle down. Occasionally it has been found that one element lags behind the others in rate of convergence. The following example will show how to deal with this.

IV. WORKED EXAMPLE

The matrix A (Table I) is a dispersion matrix of order 6×6, and is typical of those that appear in psychological research. Since it is symmetrical one-half only need be dealt with, provided checks are used.

TABLE I. DISPERSION MATRIX A

11·26	3·27	5·39	2·38	2·91	2·62
3·27	26·46	11·45	3·96	6·29	11·58
5·39	11·45	25·79	4·21	7·35	16·40
2·38	3·96	4·21	5·14	2·22	1·66
2·91	6·29	7·35	2·22	7·45	5·45
2·62	11·58	16·40	1·66	5·45	26·58

The work-sheet for the first set of computations is shown in Table II. Three lines are allowed for each row. The matrix is written down in the first line of each row, omitting the lower left-hand half and the leading diagonal. The sums are entered at the end. The matrix is then bordered, at the top and left, with the reciprocals of the diagonal elements, found from, say, Barlow's *Tables*. Using the reciprocal as a constant multiplier in the machine, we multiply each element and record the result beneath it in the second line, e.g., in the fourth row, $\cdot196 \times 2{\cdot}22 = \cdot44$, etc. The product of the reciprocal and the sum of the row are entered in the check column, $\cdot196 \times 3{\cdot}88 = \cdot76$, and compared

with the sum for the second line, ·77. The products (in the second lines) are now summed in columns ; and the sums entered at the bottom of the table, e.g., in the second column, ·47 + ·44 = ·91. Each element is next multiplied by the reciprocal at the top of the corresponding column, and the product placed in the third line, e.g., ·039 × ·47 = ·02, etc. The product of the reciprocal at the top with the sum at the bottom of the column is entered in the check row, e.g., ·039 × ·91 = ·04. The columns are then summed and compared with the checks. We now reverse the signs of the figures in the third lines, and complete the rest of the matrix. (In the table, the reversal of the signs is shown only in the lower left-hand portion of the matrix.) This matrix is C_1 unmodified and without its leading diagonal. Where A is a correlation matrix, it will be identical with A, except that the signs of the elements are reversed and the leading diagonal is omitted. It will be convenient to write it out on the left half of a fresh sheet of paper.

TABLE II. WORK-SHEET

	·088	·038	·039	·196	·135	·038	Σ	Check
	—	3·27	5·39	2·38	2·91	2·62	16·57	
·088	—	·29	·47	·21	·26	·23	1·46	1·46
	—	·01	·02	·04	·04	·01		
		—	11·45	3·96	6·29	11·58	33·28	
·038		—	·44	·15	·24	·44	1·27	1·26
	−·01	—	·02	·03	·03	·02		
			—	4·21	7·35	16·40	27·96	
·039			—	·16	·29	·64	1·09	1·09
	−·02	−·02	—	·03	·04	·02		
				—	2·22	1·66	3·88	
·196				—	·44	·33	·77	·76
	−·04	−·03	−·03	—	·06	·01		
					—	5·45	5·45	
·135					—	·74	·74	·74
	−·04	−·03	−·04	−·06	—	·03		
						—		
·038						—		
	−·01	−·02	−·02	−·01	−·03	—		
Σ_2		·29	·91	·52	1·23	2·38		
Σ_3		·01	·04	·10	·17	·09		
Check		·01	·04	·10	·17	·09		

The next step is to find the elements in the leading diagonal in their final modified state. This is done by multiplying the first row of A into the first column of C_1, omitting the first element. For this stage the elements of A may be conveniently replaced by whole numbers. The customary computer's trick is to copy the first row of A on a sheet of paper, writing the elements in a column against the right-hand edge. This should be clearly labelled. It is then placed against the left side of the first column of C_1, so that the corresponding elements are adjacent. It is now a simple matter to multiply adjacent elements, summing the results on the machine, and clearing out the multiplier each time. We do this for the second to the sixth elements, e.g.,

$$(3 \times -\cdot 01) + (5 \times -\cdot 02) + (2 \times -\cdot 04) + (3 \times -\cdot 04) + (3 \times -\cdot 01) = -\cdot 36.$$

We now put the first element of A into the machine as multiplicand, and multiply it until the total result is equal to *one* or just over (in this case 1·2). The multiplier that produced this (·14) is the first element of C_1, and is written down in its appropriate place. The total product, which forms the first element of the first column of AC_1, is then written down in its place on the right half of the page.

The next stage repeats the process for the second element on the second column of C_1 and AC_1. The second row of A is written as a column and placed against the second column of C_1 ; the pairs of elements are multiplied together and summed, then the second element of the A column is put into the machine and multiplied by hand until the final product is *one* or just over. The multiplier

and the product are then written down as the second elements in the leading diagonals of C_1 and AC_1.

When all the diagonal elements of C_1 and AC_1 have been computed, the rest of the multiplication of A with C_1 is completed. The slip with the first row of A is placed against the second column of C_1, and multiplied out ; then against the third column ; and so on, until all the elements of the first row of AC_1 have been filled in. This is repeated with the other rows. The result is seen on the right side of Table III.[1] At first sight the results do not look promising : in spite of overbuilding in the first and fifth columns, all show large negative elements ; only two elements, in the third and last column, have managed to swing past the zero. Moreover, as usually happens, the matrix thus obtained is no longer symmetrical.

TABLE III

C_1 modified : i.e., $(C_1 + X)$						$A(C_1 + X)$					
(·14)	−·01	−·02	−·04	−·04	−·01	1·22	−·20	−·10	−·17	−·09	−·21
−·01	(·07)	−·02	−·03	−·03	−·02	−·50	1·05	−·28	−·53	−·24	−·21
−·02	−·02	(·08)	−·03	−·04	−·02	−·43	−·44	·97	−·63	−·30	·09
−·04	−·03	−·03	(·31)	−·06	−·01	−·08	−·08	−·07	1·10	−·12	−·19
−·04	−·03	−·04	−·06	(·27)	−·03	−·16	−·13	−·07	−·34	1·14	−·17
−·01	−·02	−·02	−·01	−·03	(·07)	−·57	−·29	·23	−1·02	−·51	1·09

TABLE IV

C_2 modified						AC_2 modified					
(·11)	·00	−·01	−·05	−·03	·01	1·00[5]	−·02	·00	−·16	−·08	·07
·00	(·05)	·00	−·03	−·03	−·01	−·03	·90	−·08	−·02	−·07	·12
−·01	·00	(·07)	−·02	−·03	−·05	·07	·06	·88	·21	·10	−·08
−·05	−·03	−·02	(·27)	−·04	·03	−·09	−·04	·02	1·00	−·06	·04
−·03	−·03	−·03	−·04	(·21)	·00	−·03	−·03	−·06	−·02	·98	·05
·01	−·01	−·05	·03	·00	(·07)	·14	·10	−·40	·22	·16	1·00

TABLE V

C_3 modified						AC_3 modified					
(·106)	−·001	−·010	−·032	−·025	·005	1·000	−·014	·052	−·026	−·009	·040
−·001	(·055)	−·001	−·027	−·027	−·013	−·015	1·006	·112	·007	−·001	·084
−·010	−·001	(·074)	−·024	−·028	−·035	·075	·061	·963	·154	·138	·098
−·032	−·026	−·024	(·256)	−·034	·026	−·006	−·006	·040	·999	−·006	·042
−·026	−·027	−·028	−·034	(·208)	−·005	−·003	·001	·056	·017	·998	·045
·006	−·014	−·035	·026	−·005	(·064)	·067	·055	·053	·141	·107	1·006

The next stage follows the iterative routine described above. AC_1 is subtracted from $2I$, and the remainder pre-multiplied by C_1, giving the new approximation for the inverse. The new inverse should be made symmetrical by replacing the differing corresponding elements by their mean. After this, the missing leading diagonal is filled in as already described ; and the multiplication of AC_2 is completed (Table IV). The new inverse plainly shows a great improvement. The result of the next iteration (using the full values of A) is shown in Table V. In spite of working with only two decimal places, the large errors have already gone. Two more cycles would increase the accuracy to about five decimal places.

[1] The leading element in the matrix on the right of Table III is produced as follows :—

$$11{\cdot}26\times{\cdot}14 + 3\times(-\ {\cdot}01) + 5\times(-\ {\cdot}02) + 2\times(-\ {\cdot}04) + 3\times(-{\cdot}04) + 3\times(-\ {\cdot}01)$$

In multiplying the largest figure (·14) it seems wiser to retain *two* decimal places in the multiplier.

Very rarely it may happen that one column, and its corresponding row, lags behind the others in its convergence. The method of dealing with such a column may be illustrated here. The third column and row in AC_3 contain large errors, chiefly positive, though the element in the diagonal is low ; and may therefore serve as an example. Since the third column is the product of A with the third column of C_3, the negative elements of that column are plainly too low. They must be increased ; but which elements, and by how much ? The solution depends on the symmetry. The first element of the third column of C_3 must be altered so that, when the first element of the third row of C_3 is correspondingly changed, the first element of the third row of AC_3 is brought to zero. This is done by building up as before : the inner product of the third row of A and the first column of C_3 is obtained on the machine, leaving to the last the product of the element to be modified. It is then built up to make the final sum as close as possible to zero. The first column of AC_3 must then be recalculated. Since we know by how much we have altered the element in C_3, we can compute the corrections required for altering the elements of the whole column.

TABLE VI

C_3 adjusted					AC_3 adjusted					
·106	−·001	(−·012)	−·032	−·025	·005	·989	−·014	·025	−·041	−·021	·040
−·001	·055	−·001	−·027	−·027	−·013	−·038	1·006	·100	−·070	−·027	·084
(−·012)	−·001	(·078)	(−·029)	(−·033)	−·035	·023	·061	·997	−·046	·046	·098
−·032	−·027	(−·029)	(·261)	−·034	·026	−·014	−·006	·015	1·004	−·016	·042
−·026	−·027	(−·033)	−·034	(·213)	−·005	−·018	·001	·031	−·009	·998	·045
·006	−·014	−·035	·026	−·005	·064	·034	·055	·078	·067	·052	1·006

TABLE VII

C_4					AC_4					
·1063	·0007	−·0137	−·0291	−·0230	·0043	1·0005	−·0012	·0033	·0091	−·0045	−·0001
·0007	·0556	−·0041	−·0234	−·0253	−·0151	−·0009	·9999	·0041	·0135	−·0100	−·0007
−·0137	−·0041	·0836	−·0308	−·0356	−·0392	·0066	·0025	1·0010	·0006	−·0029	−·0012
−·0291	−·0234	−·0308	·2579	−·0322	·0226	·0046	·0031	·0007	1·0000	−·0059	−·0006
−·0230	−·0253	−·0356	−·0322	·2142	−·0067	·0005	−·0011	·0016	−·0024	1·0001	−·0003
·0043	−·0151	−·0392	·0226	−·0067	·0679	·0026	·0004	·0006	·0010	−·0012	·9993

When we turn to the second element of the faulty column and row, we see that the second element in the row is nearly zero. Hence, if we use this method, there will be little change. The element in the row is large and is therefore recalculated, building up the appropriate element of C_3 to bring the final result in AC_3 close to zero. As it is negative and has been made larger, this will add a negative quantity to each of the elements in the fifth column of AC_3, and in particular to the element which lies in the leading diagonal. This element must also be rebuilt. That in turn will add a positive quantity to the whole column. Hence it will be advisable, when altering the third element of column five of C_3, to see that the corresponding element of AC_3 is swung to the negative side. With this value settled, the fifth element of column five in C_3 is built up to make the corresponding element of AC_3 equal to *one*. The same is done for column six of C_3, altering the third and the sixth elements. Three new values have been found in the third row of C_3 ; and they are substituted in the third column. The third element is now rebuilt. The fourth element in the sixth row of AC_3 is disproportionately large, and therefore has also to be reconstructed. Finally, the rest of the modified matrix, AC_3, is recalculated. The result is in Table VI.

The greatest difference between I and AC_3 is now 0·1. A further iteration cycle gives Table VII. At two-decimal accuracy, the check product has only five elements which are not perfect, and the maximum error is only 0·01. At three-decimal accuracy, the approximation to a unit matrix is quite good ; and, if that is all that is required, it would be worth adjusting the five largest elements. It is interesting to compare C_4 with C_2. When it is remembered that C_1 is a simple modification of the original matrix, and, in the case of a correlation matrix, involves merely a change in signs, it is astonishing how good C_2 has proved to be. In general, the bigger the original matrix the better the result in C_2.

It is not easy to decide how this method compares with others in speed. Where the original variables are not highly correlated, it is certainly quicker, because of the rapid convergence. It is therefore particularly useful where the variables are derived variables, e.g., canonical variates, factor-measurements, etc., known to be (more or less) uncorrelated. One factor, which favours such iterative methods but slows down direct methods, is well known, but difficult to estimate in practice. In most computations, the usual check is the ' sum check ' at the end of the row or column. When this shows that a mistake has been made, the elements of the row have to be recalculated until it is found—a laborious procedure when a large number of figures is necessary. The commonest mistake is in the sign of one of the products ; and this may often be repeated each time. The computer knows that one of the elements is wrong ; but may recalculate the row without detecting where it lies. In the iterative method, the changes that occur from one iteration to the next are always small and tend to be confined to the last of the significant digits. If a mistake appears, it can easily be detected by comparing the elements with those in the previous iteration. This check is so much more efficient than the usual sum checks that many omit the latter.

V. SUMMARY

The advantages of iterative methods of computation over direct methods are summarized, with particular reference to the inversion of matrices. The main part of the paper demonstrates an extremely simple method for finding such an approximation, and then accelerating its convergence. A worked example is given, deliberately chosen to illustrate the chief difficulties.

REFERENCES

1. Hotelling, H. (1943). ' Some new methods in matrix calculation.' *Annals of Math. Stat.*, XIV, 1–34.
2. Hotelling, H. (1946). ' Practical problems of matrix calculation.' *Proc. Berkeley Symposium.*
3. Frazer, R. A., Duncan, W. J., and Collar, A. R. (1938). *Elementary Matrices.* Cambridge : University Press.
4. Richardson, L. F. (1950). ' A purification method for computing the latent columns.' *Phil. Trans. Roy. Soc.*, Ser. A., CCXLII, 439–91.
5. Burt, C. (1951). ' The numerical solution of linear equations.' *Brit. J. Psych., Stat. Sect.*, IV, 31–54.

Vol. V
Part III

The British Journal of Psychology,
Statistical Section

November,
1952

AN INVESTIGATION INTO THE PROGNOSTIC VALUE OF CERTAIN ARITHMETIC TESTS AT THE AGE OF ELEVEN PLUS

By JOHN SUTHERLAND
Moray House, University of Edinburgh

I. *Introduction and Design of Experiment.* II. *The Tests.* III. *Results of Tests.* IV. *The Follow-Up.* V. *Summary.*

I. INTRODUCTION AND DESIGN OF EXPERIMENT

In a large number of education areas attempts are made to predict the ability of a child to benefit by secondary education by means of three tests applied at the age of 11 +, namely, tests of Intelligence, English, and Arithmetic, usually of the short answer type ; and in recent journals there has of late been much discussion of the value of the results. This paper deals with one test only—that for Arithmetic. The standard form of the Moray House test, as used for this purpose, comprises two sections with 50 items in each. The first is concerned with mechanical arithmetic (the four rules and the use of elementary tables of money, length, weight, capacity, and time). The second consists of problems most of which may be solved mentally. Of these a few require a knowledge of vulgar fractions. A quarter of an hour is allowed for each section.[1]

The object of the investigation described in the following paper was to discover, first, whether the present form is the best that can be devised, and, secondly, whether the present allowance of time is the most suitable. The best form of test is taken to be that which correlates most highly with success in the secondary school ; and accordingly a follow-up of all the pupils tested in the initial phase of the enquiry was subsequently carried out. The experiment was designed in a way suited to an analysis of variance. Thus there were four tests, I, II, III, IV ; four groups of children, A, B, C, D ; four times, $\frac{1}{2}$ hour, $\frac{3}{4}$ hour, 1 hour, and $1\frac{1}{2}$ hours ; and four days, α, β, γ, δ. Each group of children took each test and each period ; each test was given on each day, with each allowance of time, and on an equal number of occasions. The advantage of this scheme is that tests, periods, groups, and days are all mutually orthogonal.

II. THE TESTS

For the purposes of the present experiment four tests were constructed. The first consisted of 100 problems in mechanical arithmetic like those in the first half of the Moray House test ; the second of 100 problems like those in the second half ; the third consisted of two sections, viz., (*a*) 50 items in mechanical arithmetic, but including simple proportion, profit and loss, and bills, (*b*) 50 items containing rather more problems in proportion, profit and loss, and decimal fractions than the standard

[1]As a result of the investigation here described, the time allowed has since been extended to 20 minutes, and items involving a knowledge of profit and loss have been included.

test ; the fourth test was a standard Moray House test consisting of 100 items as above. All four tests were tried out in Moray House Demonstration School in a pilot experiment in June, 1941. On the basis of the answer patterns certain items were rejected and others rearranged in order of difficulty. The results indicated that one and a half hours was as good as unlimited time.

In the main experiment nearly 800 children were tested, divided into four groups, containing approximately the same number of pupils, and of approximately the same level as regards intelligence and school ability. Since it was not feasible to test the entire 11-year-old population of Edinburgh, a representative sample was selected. To facilitate the follow-up the tests were applied in the first week of the new autumn term when the pupils had settled down in the secondary schools, but had not yet received sufficient secondary education to influence their scores. The tests were administered on four successive days to the whole first year in a 5-year secondary school (senior secondary) and in two 3-year secondary schools (junior secondary), each the largest of its type. Each school started work at the same time on each day.

III. RESULTS OF TESTS

Owing to absence and other reasons only 646 children sat for all four tests.

Distribution of Age. Their mean age was 145.2 mos. Thus, when sitting for the Transfer Examination in May their age would have been approximately 141.7 mos. The actual age of the whole group of 3,288 children tested in May, 1941, averaged 144.0 mos., and of the 2,040 examined in December, 1940, 145.5 mos. The ages of the individual pupils ranged from 10 yrs. 3 mos. to 14 yrs. 4 mos. ; but over 92 per cent. fell between the ages of 11 yrs. 6 mos. and 12 yrs. 10 mos., and the standard deviation was only 5.1 mos.

Distribution of Intelligence. The mean I.Q. was 105·10, and the standard deviation was 12·64. For those tested in May the figures were 100·52 and 13·86, and for those tested in December 94·15 and 13·04. The fact that the experimental group is slightly younger, more intelligent, and more homogeneous than the whole-year group is probably due to two main causes : (i) our group is a sample of the qualifying candidates who have done well enough to be admitted to a secondary school (senior or junior) ; (ii) the proportion of senior secondary school pupils (usually the most intelligent) is a little too high, namely, 37·8 per cent. But these slight deviations are unlikely to affect the validity of the results.

Reliability. Reliability coefficients were calculated by the split half method for random samples of 50 pupils taken from each group (Table I).

TABLE I. RELIABILITY COEFFICIENTS

Group	Test 1	Test 2	Test 3	Test 4	½ hour	¾ hour	1 hour	1½ hours
A	·911	·913	·899	·871	·899	·911	·871	·913
B	·955	·883	·947	·960	·955	·947	·883	·960
C	·891	·943	·911	·948	·943	·948	·891	·911
D	·966	·943	·952	·952	·952	·943	·952	·966
Mean	·938	·925	·931	·940	·940	·939	·906	·943

The means were obtained by using the z-transformation and transforming the results by the Spearman-Brown formula. The tests appear to be of about equal reliability ; and it is doubtful whether the difference shown by the 1-hour period has any significance, since the coefficient rises again for the 1½-hour period.

Analysis of Variance. For simplicity in calculation the results were arranged in the two squares shown in Tables II and III.

TABLE II. MEANS FOR TESTS AND TIMES

Test	Time					
	½ hour	¾ hour	1 hour	1½ hours	Total	Mean
1	58·73	77·48	72·76	67·23	276·20	69·05
2	52·42	61·56	68·76	82·23	264·97	66·24
3	64·37	58·69	63·78	70·32	257·16	64·29
4	63·12	66·51	83·30	78·02	290·95	72·74
Total	238·64	264·24	288·60	297·80	1,089·28	
Mean	59·66	66·06	72·15	74·45		68·08

TABLE III. MEANS FOR DAYS AND GROUPS

Day	Group					
	A(134)	B(175)	C(168)	D(169)	Total	Mean
α	83·30	58·69	52·42	67·23	261·64	65·41
β	82·23	58·73	66·51	63·78	271·25	67·81
γ	77·48	68·76	70·32	63·12	279·68	69·92
δ	64·37	78·02	72·76	61·56	276·71	69·17
Total	307·38	264·20	262·01	255·69	1,089·28	
Mean	76·85	66·05	65·50	63·92		68·08

The figures in the two tables are exactly the same, but entered in a different order. Each figure in the body of the table is the mean of from 134 to 175 individual scores. Owing to a misunderstanding, the results from one class in Group A had to be rejected. As this was the bottom class in the school for the year in question, this group is therefore the smallest, the best, and the most homogeneous, as indeed is shown by the results obtained.

Fisher's analysis of variance ((2), pp. 35–43) was applied to the values in those tables; and the results are shown in Table IV. (G.L. denotes 'Græco-Latin Square.')

TABLE IV. ANALYSIS OF VARIANCE (A)

	Degrees of Freedom	Sum of Squares	Mean Square	Error Mean Square
Tests	3	161·560	53·853	10·339
Groups	3	419·740	139·913	26·862
Times	3	528·474	176·158	33·821
Days	3	44·409	14·803	2·842
Remainder of G.L. (Error) ..	3	15·617	5·209	
Total for G.L.	15	1,169·800		

The difference between tests is significant at the 5 per cent. level ; between groups, at the 1 per cent. level ; times, at the 1 per cent. level ; days, at the 20 per cent. level.

This method, however, does not utilize all the information. It makes no use of the fact that there are almost 200 children in each group, and does not allow for the fact that, owing to the small size of one group, the rows and columns are not strictly orthogonal. A more comprehensive method was therefore employed.

TABLE V. ANALYSIS OF VARIANCE (B)

	Degrees of Freedom	Sum of Squares	Error Mean Square
Græco-Latin Square	15	179·745	
Between children in Group A	133	55·782	
Between children in Group B	174	213·749	
Between children in Group C	167	156·845	
Between children in Group D	168	231·548	
Error	1,926	85·679	44·48
Total	2,583	923·348	
Total calculated independently		919·675	

In the first place, a more accurate and independent method of finding the error variance was used. The results obtained are shown in Table V. The difference (3,673) between the two values for the total sum of squares is accounted for by certain approximations in the calculations. Since the error mean square of 44·48 is based on 1,926 degrees of freedom instead of 3, it provides a far more accurate estimate. In the second place we want to allow for the fact that, owing to the reductions in Group A, the rows and columns of the Græco-Latin Square are not strictly orthogonal. Following Yates's method ((1), pp. 108–45), we obtained the results shown in Table VI.

TABLE VI. ANALYSIS OF VARIANCE

	Degrees of Freedom	Sum of Squares	Mean Square	Error Mean Square
Tests	3	25,800·63	8,600·21	193·35
Groups ..	3	75,915·00	25,305·00	568·91
Times	3	84,398·13	28,132·71	632·48
Days	3	7,100·34	2,366·78	53·21
Error	1,926	85,679·00	44·48	

For significance at the 0·1 per cent. level (the highest value quoted in the table) the variance ratio needs to be only 5·42 ((2), pp. 35–43). Every difference is therefore highly significant.

We have now to find which tests, times, and groups differ significantly from the rest. This was found in the usual way by calculating values of $t = \frac{D}{\sigma_D}$, where D is the difference between the means of the tests, etc., and σ_D is the standard deviation of the differences (derived from the error mean square in Table V).

The means and standard deviations for each test, time, and day are shown in detail in Tables VII and VIII. The letters of the groups have been entered above the scores to avoid confusion due to the inequality in the performance of the groups. The differences and critical ratios are shown in Table IX.

TABLE VII. MEANS AND STANDARD DEVIATIONS

Test	Time									
	½ hour		¾ hour		1 hour		1½ hours		Aggregate	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	B 58·73	15·93	A 77·48	10·76	C 72·76	13·13	D 67·23	18·71	68·48	15·13
2	C 52·42	18·95	D 61·56	20·49	B 68·76	21·10	A 82·23	12·18	65·42	18·84
3	A 64·37	11·03	B 58·69	17·24	D 63·78	18·43	C 70·32	15·98	64·22	16·16
4	D 63·12	19·36	C 66·51	17·74	A 83·30	12·03	B 78·02	18·46	72·22	17·40
	59·41	16·90	65·37	17·22	71·51	16·86	74·07	16·77	67·59	16·88

TABLE VIII. MEANS AND STANDARD DEVIATIONS

Test	Tuesday		Wednesday		Thursday		Friday		Aggregate	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	D 67·23	18·71	B 58·73	15·93	A 77·48	10·76	C 72·76	13·13	68·48	15·13
2	C 52·42	18·95	A 82·23	12·18	B 68·76	21·10	D 61·56	20·49	65·42	18·84
3	B 58·69	17·24	D 63·78	18·46	C 70·32	15·98	A 64·37	11·03	64·22	16·16
4	A 83·30	12·03	C 66·51	17·74	D 63·12	19·36	B 79·02	18·43	72·22	17·40
Mean	64·40	17·19	66·95	16·44	69·50	17·59	69·51	16·49	67·59	16·88

The probability that the differences are due to chance is less than 0·001 in every case except the last (1 hour and 1½ hours, probability 0·07). Thus, with this last exception, all the tests and all the times give significantly different results. We note that no advantage is gained by increasing the allowance of time to 1½ hours. Group A alone differs significantly from the others, owing doubtless to the circumstance mentioned above. The mean score increases fairly steadily as the time increases up to 1 hour, but far less for the last ½ hour; 1 hour was apparently enough for most pupils to do all they could. The mean for each day shows a rise up to the third day, and remains almost steady after that. Apparently training or practice had reached its maximum effect by then. It may be noted that the difference between days was much smaller than the difference due to any other source. Between the means for the first and second days the difference is 2·55 points, and between the second and third days 2·55 points. On a standard scale, with $\sigma = 15$, it will be slightly less. Though there was only one day between the tests, these values seem on the whole to be rather less than most

TABLE IX. SIGNIFICANCE OF DIFFERENCES

Tests	Difference Between Means (D)	D/σ_D	Times	Difference Between Means (D)	D/σ_D
1 and 2	2·96	7·977	½ hr. and ¾ hr.	6·58	17·73
1 and 3	4·89	13·179	½ hr. and 1 hr.	13·49	36·36
1 and 4	4·78	12·882	½ hr. and 1½ hrs.	14·57	39·27
2 and 3	1·93	5·201	¾ hr. and 1 hr.	6·91	18·62
2 and 4	7·74	20·859	¾ hr. and 1½ hrs.	7·99	21·53
3 and 4	9·67	26·061	1 hr. and 1½ hrs.	1·08	2·911

of those quoted by Vernon (3) and (4) ; but they are of the same order. Three months earlier the children tested had sat for a transfer examination including tests of the three types mentioned above. To this extent they may therefore be described as ' sophisticated ' before taking the first test in the present experiment. The fact that after the third testing there is no further increase agrees with Vernon's findings.

IV. THE FOLLOW-UP

During the three years succeeding the tests, a follow-up was carried out in the three schools. Marks were obtained for all school subjects : in some cases there were three examinations per year, in others two. The unweighted average for the year was used. After considering the requirements of the examinations and the opinions of headmasters, it was decided to define " success in the secondary school " by the quotient $X = (2E + 2M + s + a + l)/7$, where $E = (2e + g + h)/4$ and $M = (ar + al + g)/3$, and the small letters denote English, Geography, History, Arithmetic, Algebra, Geometry, Science, Art, and Languages (the average mark for the year being expressed as a percentage).

Owing to children leaving school there was considerable wastage. Of the 646 who originally sat for all four tests, 454 completed the first two years of the secondary school course and 188 completed the whole three years. The consequent reduction in the spread of ability in the sample is illustrated in Table X.

TABLE X. REDUCTION IN DISPERSION DUE TO SELECTION

Test	I.Q.	Arith. 1	Arith. 2	Arith. 3	Arith. 4
Whole Group (646)	12·6	15·1	18·8	16·2	17·4
After 3 Years (188).. ..	10·4	9·7	9·8	10·4	9·5

The correlations of each of the arithmetic tests with X and M at the end of the second and third year course were calculated for each group separately. Average correlations were derived for each of the tests, and each of the four periods, with X and M. The values are shown in Table XI, together with the correlations with the Intelligence Quotient and with the class teachers' scaled arithmetic estimates. Since the reduction in scatter is substantially the same for every test, no correction has been made for selection.

TABLE XI. CORRELATIONS OF ARITHMETIC TESTS WITH SUCCESS IN THE SECONDARY SCHOOL

Test	School Average (X)		Mathematics (M)	
	Second Year	Third Year	Second Year	Third Year
1. Mechanical Arithmetic ..	·3714	·2599	·4462	·2849
2. Problem Arithmetic	·5456	·5126	·5348	·5241
3. New Test (Mech. + Prob.) ..	·4854	·4382	·5469	·4478
4. Standard M.H. Arith. Test ..	·4960	·4446	·5511	·4244
Intelligence Quotient	·5779	·5226	·5562	·4581
Class Teachers' Arith. Est. ..	·5264	·4670	·5657	·4428

The test which correlates most highly with both general school success and mathematical ability at the second-year level and with mathematical ability at the third-year level is Test 2. This was composed entirely of problems. It was compared with intelligence and the teachers' arithmetic estimate as a means of predicting secondary school success. The regressions on secondary school success were as follows: Problems, ·5456; Intelligence, ·6500; Teachers' Estimate, ·2179. Thus even in a battery including an intelligence test the problem arithmetic test still contributes almost as much to the prediction as the intelligence test and far more than the teachers' estimate. The results obtained with the new test (Test 3) agree fairly well with those from a standard Moray House Arithmetic Test (Test 4). That was to be expected, since only a small number of new items were included. Detailed answer patterns for Tests 3 and 4 showed that, while fewer children succeeded in getting the new type of items correct, these did not discriminate between the children better than the old. There is therefore no case for their inclusion apart from the desirability of covering the whole syllabus in the examination.

It was also decided to determine which period gave results correlating most closely with success. These coefficients are shown in Table XII.

TABLE XII. CORRELATION BETWEEN SCORES WITH VARIOUS TIME ALLOWANCES AND SUCCESS IN SECONDARY SCHOOL

Time	School Average		Mathematics	
	Second Year	Third Year	Second Year	Third Year
½ hour	·4629	·4070	·4688	·4390
¾ hour	·4494	·3610	·5139	·3601
1 hour	·4839	·4219	·5241	·4252
1½ hours	·5139	·4777	·5744	·4676

The predictive value seems to increase with the length of the time allowance. For the third year the standard errors vary from ± ·063 for the smallest difference to ± ·056 for the largest. Hence none of the differences is statistically significant; and in any case it is doubtful whether they are large enough to justify an authority in allowing as much as 1 hour or 1½ hours. Suppose that we give the children two ½-hour tests of arithmetic, each of which correlates with third-year success to the extent of 0·4 and with each other to the extent of 0·9. On combining the results (with equal weighting) we should get a correlation with success amounting to 0·42. Thus very little increase in predictive value is obtained by adding more *similar* items (which is in effect what happens when the time is increased). Nor does increase of

time increase reliability (cf. Table I). Thus a $\frac{1}{2}$-hour test, properly constructed, will assess the child's arithmetical ability reasonably well ; and any extra time available would be more profitably spent in administering a different test which correlates highly with secondary school success and *as low as possible with the other tests.*

V. SUMMARY

1. Four different arithmetic tests, a standard Moray House test, a modified Moray House test, a test of problem arithmetic, and a test of mechanical arithmetic, were administered to 646 children in the first week of the secondary school course. Four different time allowances were used, and the tests applied on four successive days. A follow-up was carried out during the succeeding three years.

2. An analysis of variance indicated that the differences between tests, between times, and between days were all significant. The mean score increased with the time allowed up to 1 hour ; it also increased on successive days up to the third day. This agrees with the results of other investigators, though the amount of the increase here obtained is not so great as has been reported by other writers.

3. The follow-up showed that the problem arithmetic test was the best predictor of secondary school success, and that the problem test was almost as successful as the intelligence test, and much superior to the teachers' estimate. The correlations with secondary school success increased with increase in the time allowance, but the differences were not significant. It was concluded that an increase of time for a test consisting of similar items does little to improve its predictive capacity.

REFERENCES

1. Yates, F. (1933). ' The Principles of Orthogonality and Confounding in Replicated Experiments.' *J. Agric. Sci.*, XXIII, 108–45.
2. Fisher, R. A., and Yates, F. (1949). *Statistical Tables for Biological, Agricultural, and Medical Research.*
3. Vernon, P. E. (1951). ' Coaching and Practice Effects in Intelligence Tests.' *Proceedings of Psychological Congress, Stockholm.*
4. Vernon, P. E. (1952). ' Intelligence Testing : Effects of Practice and Coaching.' *Times Educ. Supp.*, January 25th, February 1st.

NOTES AND CORRESPONDENCE

THE ENTROPY OF AN ENSEMBLE WHICH IS DIVIDED INTO CLASSES

By W. E. HICK

Medical Research Council Applied Psychology Unit, Cambridge

In applications of Communication Theory, it is sometimes useful to know the total entropy of an ensemble when firstly a class is selected (with probability $p(i)$), and then one member of the class (with probability $p_i(j)$). It might be thought sufficient to take the average, weighted by the $p(i)$, of the intra-class entropies ; but this fails to take account of the uncertainty as to which class will be chosen. The correct total is the joint entropy $H(x, y)$ of the selections of class and event-within-class. Shannon's notation (1) can be used directly, if x represents the class-ensemble instead of the message-ensemble, and y is the event-ensemble (within a class) instead of the ensemble of received signals. Thus $p_i(j)$ is the chance of the jth event of the ith class, given that the ith class has occurred.

$$\begin{aligned} H(x, y) &= -\sum_{ij} p(i,j) \log p(i,j) \\ &= -\sum_{ij} p(i)p_i(j)\{\log p(i) + \log p_i(j)\} \\ &= -\sum_{i} p(i) \sum_{j} p_i(j) \log p_i(j) - \sum_{j} p_i(i) \sum_{i} p(i) \log p(i) \\ &= H_x(y) + H(x) \end{aligned}$$

$H_x(y)$ is the conditional entropy of y when x is known. It represents the amount by which the class entropy $H(x)$ fails to determine the final event, and corresponds to the disturbance or ' noise ' in a communication system. Note that the classes need not be equal ; the smaller classes can be supposed to have their deficiencies made good by fictitious events of zero probability.

1. Shannon, C., and Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana : University of Illinois Press.

BIPOLAR FACTORS AS A CAUSE OF CYCLIC OVERLAP

To the Editors, British Journal of Psychology : Statistical Section.

Sirs,—In his article on ' The Factorial Study of Temperamental Traits ' (this *Journal*, I, p. 181), Professor Burt summarizes an investigation, carried out (it seems) as early as 1915, with the object of establishing the existence of both general and special (or ' group ') factors underlying the primary emotions. Several recent publications have referred to this enquiry (e.g., H. Eysenck, *The Scientific Study of Personality*, 1952, pp. 41, 114, 300) ; and it would be helpful to know a little more about the methods and results, particularly how far they agree with, or differ from, those of later American investigators who have taken up much the same problem, such as L. L. Thurstone (*Psychometrika*, XVI, 1951, pp. 11–20).

We are told that the factorial procedure adopted was suggested in part by McDougall's conception of emotional or conative impulses as mental ' forces.' As forces, therefore, such tendencies were presumed to behave like forces in mechanics " combining in accordance with the familiar parallelogram law to produce observable resultants." The chief aim of the procedure proposed was apparently to reduce the numerous observed ' resultants ' to terms of a small set of independent hypothetical tendencies, much as in dynamics a resultant force is resolved into its rectangular components. The procedure is illustrated by a geometrical diagram (this *Journal*, I, Fig. 1, p. 190) ; but it is not made clear how a working formula was deduced from these assumptions in order to calculate the saturations on which the diagram is presumably based. Were they computed by Pearson's formulæ for finding ' lines of closest fit ' (the method of ' principal axes ' or ' weighted summation,' as it is variously termed), or by the formula for simple summation, given in Burt's book on *Educational Abilities* (1917) and subsequently used by Thurstone ?

The first factor, it is stated, could be identified, in the light of independent evidence, with ' general emotionality '—a characteristic which, it is assumed, is largely innate : this appears to be accepted by Spearman, since he tells us that Burt's conclusion confirmed Webb's earlier discovery of a new general factor, distinct from g (*Abilities of Man*, p. 348). But the unique feature of the research was apparently the discovery of what is termed ' cyclic overlap.' After the general factor had been partialled out, the table of residual or specific correlations indicated " a scheme which is not hierarchical and has not hitherto been described : the positive coefficients run in a diagonal line from the top left-hand corner (of the table) to the bottom right-hand corner, with outlying groups at the top right and bottom left ; and the negative coefficients run in lines parallel to the diagonal." As a result, " the series of emotional traits can be arranged, as it were, in a closed circle," analogous to the

familiar " circle of saturated colours " (spectral colours plus purples), shown, for example, in Wundt's model of the ' colour sphere ' (*Grundzüge*, II, p. 167, Fig. 191).

Now if I understand Burt's argument aright, his contention is that this peculiarity demonstrates the presence of two special factors over and above the general factor : (i) a factor for what Jung had called extraversion *versus* introversion, and (ii) a new factor for pleasant *versus* unpleasant emotions. But surely the mere presence of two bipolar factors cannot *alone* suffice to produce the phenomenon of ' cyclic overlap,' else that phenomenon would have been observed far more frequently in the other investigations since carried out by Burt's summational method or by Thurstone's version of it. What, then, are the further conditions ?

Possibly this question could be answered by an inspection of the matrix of factor-saturations. Unfortunately the article in this *Journal* reproduces only the diagram, not the figures. Holzinger and Harman have reprinted and re-analysed the correlations ; but their version is limited to 8 of the 11 traits only ; and (possibly for this reason) yields no sign of cyclic overlap (*Factor Analysis*, 1941, Tables 7.6 and 7.7). As it is difficult for some of us, particularly since the recent war, to gain access to English reports published as far back as 1915, it would be helpful to have the complete set of correlations and factor-saturations, if these are still available. The results would seem to be of some historical interest, since (so far as one can discover) they represent the earliest factorial study demonstrating the existence of ' bipolar factors.' This is a type of factor which both Spearman and Thurstone apparently deny or at least ignore ; on the other hand, Holzinger and Harman accept it, and explicitly attribute the conception to Burt (p. 104). However, my chief interest is in the relation between these factors and the circular overlap they are said to produce.

My questions are prompted by a small study I am making of (1) psychoneurotic and (2) psychotic cases. I agree with Professor Burt that a pathological disorder connotes a qualitative rather than a quantitative difference. Yet, when he himself undertakes a factor analysis of such cases, it is surprising to find that, even in his latest research (pp. 193f.), he bases that analysis on measurements for graded *variables* instead of on frequencies for the presence or absence of *qualities*. If ' factors are principles of classification,' surely a factorial procedure should be equally possible whether the classification is based on all-or-none differences of kind or on measurable differences of degree. In that case, however, is it still legitimate to look for cyclic overlap, and can a bipolar method be used ? —Yours, etc., HENRI LECLÈRE.

Reply.—Many of M. Leclère's questions have, I think, been answered, at least by implication, in more recent articles in this *Journal* (cf. II, pp. 125f., pp. 151f., III, pp. 166f.). However, several points that he has raised would seem to be of current interest to investigators engaged on problems similar to his own, and therefore perhaps deserve a fuller reply.

The early investigation to which he refers was originally described in my paper on ' General and Specific Factors Underlying the Primary Emotions ' (*Brit. Ass. Ann. Rep.*, 1915, pp. 695f.), and was based on assessments collected between 1909 and 1914 for five groups of children and adults. Space will not permit me to reproduce all the tables in detail [1] ; but I give below the two most complete, which virtually sum up the main enquiry. Table I shows the averages of the coefficients obtained by pooling the four largest groups, numbering 287 children in all : only those individuals are included who had been assessed for all [2] the traits. Table II gives the factor-saturations. In each case the order of the traits corresponds with that in the two tables printed in the original report.

[1] In the printed abstract only two tables were given (for adults and children respectively). As it had to be prepared well in advance, the figures included were those obtained in the earlier stages of the investigation only. At the meeting a complete set of tables was circulated from which the above are taken : however, the figures got from the larger group show no essential differences. The data were obtained chiefly from school children. Dr. Webb (a colleague of mine under the London County Council) decided to carry out a similar investigation with adult students at a Council Training College. His preliminary account was published in the same *Annual Report* (p. 694), but was limited to a discussion of the methods proposed ; his statistical analysis was not completed or published until later. In stating that my results " confirmed " those of Webb, Spearman, I fancy, must have really been thinking of a later paper in which I compared the two. Webb and I did not go so far as to identify the two general factors. Indeed, we both agreed that, as found in adults, the general factor appeared ' far more complex and far less dependent upon innate temperament.' At Spearman's request a fuller account of my own investigation was given in his journal (*Character and Personality*, VII, pp. 238f.) to which M. Leclère may perhaps refer : Webb's factor is there discussed on p. 253.

[2] In the first of the tables printed in the *Brit. Ass. Ann. Rep.*, each correlation was based on the largest number of children who happened at the time of writing to have been assessed for each pair of correlated traits. For traits like Sex and Fear the teachers stated that they were unable to obtain satisfactory observations in the case of a good many of their pupils. Hence the figures for these traits were based on slightly smaller numbers. For purposes of a factor analysis, however, it is desirable that all the correlations be obtained from precisely the same individuals. Otherwise the coefficients may turn out to be formally incompatible, and their joint analysis may even lead to ' imaginary ' factors (in technical language, the matrix may no longer be ' positive-definite ').

TABLE I. CORRELATIONS BETWEEN TRAITS AND SATURATIONS FOR FIRST FACTOR

Trait	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	Total
(1) Sociability	(·998)	·821	·703	·754	·723	·732	·538	·569	·540	·294	·117	6·789
(2) Sorrow	·821	(·997)	·804	·583	·561	·443	·522	·400	·249	·487	·282	6·149
(3) Tenderness	·703	·804	(·922)	·645	·372	·316	·308	·142	·241	·401	·342	5·196
(4) Joy	·754	·583	·645	(·783)	·506	·502	·343	·306	·464	·181	—·106	5·051
(5) Wonder	·723	·561	·372	·506	(·667)	·534	·377	·547	·453	·108	—·133	4·715
(6) Elation	·732	·443	·316	·502	·534	(·697)	·469	·506	·469	·113	·018	4·799
(7) Disgust	·538	·522	·308	·343	·377	·469	(·482)	·384	·251	·227	·084	3·985
(8) Anger	·569	·400	·142	·306	·547	·506	·384	(·628)	·419	·038	—·110	3·829
(9) Sex	·540	·249	·241	·464	·453	·469	·251	·419	(·467)	—·048	—·155	3·350
(10) Fear	·294	·487	·401	·181	·108	·113	·227	·038	—·048	(·347)	·304	2·452
(11) Subjection	·117	·282	·342	—·016	—·133	·018	·084	—·110	—·155	·304	(·515)	1·248
Total	6·789	6·149	5·196	5·051	4·715	4·799	3·985	3·829	3·350	2·452	1·248	47·563
Saturations	·984	·892	·753	·732	·684	·696	·578	·555	·486	·356	·181	6·897

Methods. (*a*) *With Quantitative Data.* The saturations in this enquiry were calculated by the formula for simple summation, applied to the correlation matrix with reduced self-correlations. The self-correlations (reached by successive approximation) were computed on the principle that seven common factors[1] at most are needed to account for an 11×11 table of observed correlations.

The difference between the results obtained with my own procedure and those which would have been obtained with Pearson's turned not so much on the substitution of simple summation for weighted summation (as M. Leclère rather implies), but on the substitution of reduced self-correlations for unity in the leading diagonal. Pearson's method would require eleven factors to fit the inter-correlations between eleven traits : this would greatly magnify the size of the saturations for the first few factors, and makes no allowance for factors peculiar to each test (' specific ' factors in Spearman's sense). On the other hand, with the self-correlations reduced to eliminate ' specificity,' it was possible (as I endeavoured to show) to secure a very good fit with less than half that number of factors. With the self-correlations suggested in Table I and the saturations set out in Table II, an excellent fit is obtained with five factors only : the largest residuals then remaining are all numerically less than $\pm 0{\cdot}015$. (Seven factors should yield an *exact* fit ; but for that purpose a good many further iterations would probably be necessary, and in any case it is clear that the two additional factors would have saturations well under 0·100, and so be really negligible.) Applying the significance test described in my 1909 article it was concluded that the first three factors alone were statistically significant.

(*b*) *With Qualitative Data.* The problem of applying factorial methods to qualitative ' attributes,' as distinct from quantitative ' variables,' is, as M. Leclère implies, of special interest in the field of medicine (whether psychological or not), and came much to the fore at the outset of the war. The procedure I then suggested has since been published (this *Journal*, III, pp. 166–85), and appears to work quite well with data such as M. Leclère has in mind. The procedure itself follows as a straightforward deduction from Pearson's equations for the principal transformation when applied to a ' manifold contingency table,' such as was constructed by Yule to establish total and partial ' associations ' instead of total and partial correlations. According to the special nature of the problem and the hypotheses to be tested, various minor modifications may be desirable in actual practice. (i) With Pearson's original method the figures in the diagonal of the symmetric frequency matrix should be unity, and this entails a preliminary standardization of the figures actually observed. But, if specific factors are assumed to be present, the diagonal figures will require a different sort of adjustment, analogous to that made in reducing self-correlations or ' communalities.' (ii) A fairly satisfactory discrimination is usually furnished by the bipolar factors (such as emerge with Pearson's procedure). But a more rigorous analysis can be made by taking into account the frequencies of a higher order : (cf. *loc. cit.*, p. 182). These can be used to form an orthogonal rotation matrix for converting the bipolar factors into group factors, whose saturations will express the probabilities of the various traits occurring in the several groups. Yule's equation for independence yields a simple working formula for this purpose.[2] The result demonstrates that, here as elsewhere, the ' method of summation ' furnishes appropriate saturation coefficients. It is only fair to add that, when first put forward, such procedures were criticized as being somewhat cumbrous. Mr. Patrick Slater and others, for example, suggested alternative methods of their own which certainly deserve consideration (see, for instance, this *Journal*, II, pp. 17f., and refs.) ; but in my own experience they are equally lengthy, and lead to very much the same results.[3]

Interpretations. The early investigation referred to by M. Leclère is by no means the only one in which ' cyclic overlap ' was noted : (cf., for instance, *Distribution of Educational Abilities*, 1917, pp. 59–60, and Fig. 8). Two factorial interpretations were suggested. (1) In some cases (e.g., temporal overlap) it may plausibly be assumed that, in addition to the general or basic factor, there are a small number of more specific ' group factors,' and that these ' group factors ' overlap *en échelon*. This explanation is analogous to the Young–Helmholtz theory of primary colours, where orange is regarded

[1] The statement and formal proof of the general principle giving the minimum number of common factors needed with n tests I owed to Dr. W. F. Sheppard (at that time an Inspector of Schools).

[2] *Introduction to the Theory of Statistics*, 1910, eqn. (9), p. 57.

[3] It is encouraging to find that in a chapter on ' The Quantification of a Class of Attributes ' (in a collection of papers on *The Prediction of Personal Adjustment*, issued by the Social Science Research Council, New York, as early as 1941) Dr. Louis Guttman had already reached a formula very similar to my own, as the outcome of one of his three alternative modes of approach : (cf. also Stouffer, S. A., *et al.*, *Prediction and Measurement*, 1950, ch. IX). Apparently, however, he would not agree that the ' components ' so obtained by his procedure can be interpreted in a way analogous to that adopted for factors derived from variables ; and he would therefore probably not endorse the practical applications I carried out. But Guttman's problem was not primarily to devise factorial procedures ; and its implications accordingly require a closer study. His contributions have only recently become available on this side of the Atlantic ; but M. Leclère will find a brief reference in Guilford's recent book on *Fundamental Statistics* (pp. 371f.), together with a suggestive discussion bearing on his final question.

as due to the overlapping of red and yellow, green to that of yellow and blue, purple to that of blue and red. (2) In other cases the ' imbrication ' can be regarded as the effect of two independent bipolar factors, much as Hering proposed to analyse saturated colours in terms of two bipolar components.

TABLE II. SATURATIONS

Factor	I	II	III	IV	V
(1) Joy	·732	+·160	+·308	+·272	+·196
(2) Sex	·486	+·454	+·117	+·003	−·039
(3) Sociability	·984	+·136	+·059	+·017	−·015
(4) Elation	·696	+·315	+·041	−·292	+·113
Subtotal	—	—	+·525	·000	+·255
(5) Wonder	·684	+·337	−·194	+·157	−·088
(6) Anger	·555	+·379	−·331	−·157	−·167
Subtotal	—	+1·781	−·525	·000	−·255
(7) Disgust	·578	−·015	−·219	−·201	+·211
(8) Sorrow	·892	−·357	−·198	+·188	−·025
(9) Fear	·356	−·441	−·077	+·013	+·069
Subtotal	—	—	−·494	·000	+·255
(10) Subjection	·181	−·576	+·180	−·278	−·157
(11) Tenderness	·753	−·392	+·314	+·278	−·098
Subtotal	—	−1·781	+·494	·000	−·255

As M. Leclère rightly remarks, not any pair of bipolar factors will produce the effect observed. The essential requirement is that the saturations for any two adjacent traits (including the first and last) must yield an inner product which is positive. No matter how the traits are arranged, there will always be at least two points within each bipolar factor where positive and negative saturations come together in adjacent traits. At such a point the inner product will be negative unless the saturations for the other factor (i) happen to be both of the same sign and (ii) yield a product that is much larger than the negative product resulting from the two saturations of opposite sign. These requirements can most readily be understood by looking at a typical set (artificially constructed), such as those given below.

Tests	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Factor II	+·2	+·8	+·8	+·2	−·4	−·6	−·6	−·4
Factor III	+·7	+·1	−·1	−·7	−·5	−·3	+·3	+·5

But however the phenomenon might be explained, it seemed clear that it compelled us to recognize, in addition to the general factor, the presence of certain specific factors, either group or bipolar. It was chiefly to this conclusion that my arguments were directed, since at that date the views of Spearman and his school were opposed to the existence of specific factors other than those peculiar to each single trait or test.[1]

[1] So far as possible I sought to cast my evidence in a form which would forestall the probable objections of critics at that date. Spearman's criterion for significance insisted on the crucial figures being *five* times the probable error : hence my desire to base the correlation on the largest available number in every case. Further, in addition to giving residual correlations and saturation coefficients, I also applied the alternative statistical methods put forward by Spearman himself in his 1912 paper (based on the calculation of the correlations between columns). Pearson and his followers had objected to my quoting simple residuals (differences between observed and expected correlations) as evidence for group factors, and had contended that the proper method would be to convert the residuals into coefficients of partial correlation, before proceeding to test their significance (cf. W. Brown, *Mental Measurement*, 1912, p. 93). For this reason I also printed ' residual correlations corrected for specificity ' (by the partial correlation formula), a procedure analogous to that later adopted by Lawley in what he terms the method of maximum likelihood.

Of the two interpretations I preferred the second in the case of emotional traits, since it seemed to provide a plausible explanation of the familiar schemes of temperamental types, which (as Stern had pointed out) nearly always turned on the ' crossing of antitypes.' This raised the further question : what concrete meaning do the saturations suggest for the two bipolar factors ? For the smaller (factor III) the answer was fairly obvious : it contrasts pleasant emotions with unpleasant. In fact it seemed analogous to the component contrasting *Lust* and *Unlust* in Wundt's tridimensional theory. The larger factor (factor II) appeared roughly to correspond with Wundt's component contrasting *Erregung* with *Hemmung* (or *Beruhigung*) and with his distinction between *Stark* and *Schwach* emotional reactions in his theory of temperaments. I therefore suggested that the factor might be regarded as distinguishing sthenic or aggressive emotions from the asthenic or inhibitive. This form of the double antithesis was by no means new. It may be remembered that Hume, at the outset of his analysis of the ' passions ' (*Treatise on Human Nature*, Bk. II, pt. ii, sect. 2), suggested that they could be arranged on the periphery of a square. And what I had in mind was not Jung's later classification, as M. Leclère supposes, but the earlier classifications of Hume and Wundt.[1]

With qualitative conditions (e.g., with discrete types of mental disorder, which, particularly among children, are often combined) cyclic overlap may certainly be found ; and, when this happens, I should prefer to use overlapping group factors.

M. Leclère's question about the differences between my results and Thurstone's I can only touch upon very briefly here. If I understand Professor Thurstone rightly, the research which M. Leclère has cited was not planned to confirm or refute any explicitly formulated psychological hypothesis, but merely " to determine the number and nature of factors implied in current personality schedules." Since the traits assessed were reversed (where necessary) so as always to represent ' socially desirable forms,' the first centroid factor is simply a general factor measuring ' social desirability.' But the centroid matrix is immediately rotated [2] to secure a simple structure of positive factors only (*Psychometrika*, *loc. cit.*, pp. 14–19). For the rest I am inclined to believe that as in the case of cognitive abilities, so here the ' second order factors ' will prove to approximate much more closely to my own scheme of factors than the ' first order factors.'—CYRIL BURT.

[1] I agree that the second antithesis is not unlike the distinction between introvertive (or self-regarding) emotions and extravertive (or other-regarding) emotions (*Psychological Types*, 1920, English translation, 1924, pp. 542, 567). But (i) Jung himself did not invent the terms ' extraversion ' and ' introversion,' as M. Leclère and others appear to suppose. The terms had long been in use, particularly among mystical writers. As early as 1656, Blount's *Glossographia* gives them with definitions : (e.g., " *Extraversion*, scattering of one's thoughts upon exterior objects "). Both are also used by Wesley (*Works*, 1788, VI, p. 451). (ii) Jung's account was not published until later. (iii) If (as the original usage implies) the antithesis between intro- and extra-version refers largely to the contrast between self and other *persons*, then it would seem to correspond rather to that indicated by a third bipolar factor which has appeared in some of our analyses—a factor which includes high saturations for *both* Self-assertion *and* Self-submission as well as for Sociability and Sex, and was tentatively described as ' social maturity ' or ' interest in other persons rather than in self or things.'

[2] Thurstone's rotations are determined largely by the meaning of the factors ; mine by purely mathematical considerations. Thus, when (as in M. Leclère's study) we can secure two product-sum tables for the same tests or traits (whether derived from two samples or two orders of frequency), we can argue as follows. Let L denote the normalized bipolar factor-saturations and G the normalized group factor-saturations. Factorizing R_1 by weighted summation, we should have $R_1 = L_1V_1L'_1 = F_1F'_1$, where $V_1 = F'_1F_1$. If the *bipolar* factors are the same, then $L_1 = L_2$, and $R_1R_2 = L_1V_1L'_1 . L_2V_2L'_2 = L_1V_1V_2L'_1$, which is symmetric. If the *group* factors are the same, then $G_1 = G_2$, $R_1 = G_1D_1G'_1$, and $R_2 = G_2D_2G'_2 = G_1D'_2G_1$, where D (like V) is a diagonal matrix of factor-variances. Now let $G = F_1H$ (H being an orthogonal rotation matrix). Then, as usual, W_1, the matrix of weights for calculating the bipolar factor-measurements from test measurements, $= V_1^{-1}F_1$; and H can easily be got by factorizing the weighted correlation matrix $W'_1R_2W_1$: for $W'_1GD_2G'W_1 = V_1^{-1}F'_1 . F_1HD_2H'F'_1 . F'_1V_1^{-1} = HD_2H'$. We can thus test whether $G_1 = G_2$. Such methods have been found practicable by several of our research-students ; and it was largely on these grounds that we concluded that group factors are usually more stable than bipolar factors (cf. *Brit. J. Educ. Psychol.*, IX, p. 67, this *Journal*, III, p. 70, and refs.).

BOOK REVIEWS

Educational Measurement. Edited by E. F. Lindquist. Washington : American Council on Education. Pp. xx + 818. 1951. $6.00.

During the past thirty-three years the American Council on Education has founded or fostered numerous enterprises in the field of mental and scholastic testing. Most of its practical activities were turned over in 1948 to the Educational Testing Service, which it helped to set up ; but it still retained an interest in planning and assisting projects of a more general or theoretical nature. As soon as the war was over, its Committee on Measurement and Guidance arranged a special conference to consider and plan an authoritative volume on the subject of measurement in education. The topics to be covered were distributed among eighteen chapters ; and an authoritative writer, aided by half-a-dozen collaborators, was entrusted with the preparation of each one. The whole has been collated and edited, with admirable skill, by Dr. E. F. Lindquist, assisted by a small group of sub-editors.

The volume is divided into three main parts. Part I deals with the general purposes of measurement in education ; Part II with the methods of test-construction ; and Part III with the theory or ' philosophy ' of mental measurement. In the concluding chapters the statistical psychologist will find some of the most suggestive contributions in the book. The introductory account of the fundamental nature of measurement by Irving Lorge is clear and concise ; and is followed by more detailed discussions by various well-known authors on such topics as ' Reliability,' ' Validity,' ' Units and Norms,' and ' Batteries and Profiles.' The volume will unquestionably provide, not only a useful handbook for senior students of educational psychology, but an indispensable work of reference for investigators embarking on research in this or allied subjects. *C. B.*

Handbook of Experimental Psychology. Edited by S. S. Stevens. New York : John Wiley. Pp. xii + 1,436. 1951. £6.

" The need for a handbook," says Dr. Stevens, " is a periodic occurrence." Murchison's *Handbook of General Experimental Psychology*, published in 1934, had already dropped out of print when the recent war commenced. At one time the American Psychological Association contemplated the production of an up-to-date volume as an official venture of the Association ; but, we are told, when a publisher came forward with a concrete plan the Association " withdrew in favour of private enterprise."

The editor and his thirty-four contributors have addressed themselves " directly to the advanced scholar "—to the postgraduate student who requires a more detailed text-book and to the specialist who needs a reference book on matters outside his own particular field. Each chapter has been written by a well-known authority on the subject. Professor Langfeld is advisory editor for the series in which the volume appears ; and Professor Stevens has had further assistance from Professor Boring and Professor C. T. Morgan.

Professor Stevens begins with an introductory chapter on ' Mathematics, Measurement, and Psychophysics ' ; and the keynote of the volume is expressed in one of his opening sentences : " When description gives way to measurement, calculation can replace debate." Mathematics he treats as " a branch of logic." After a succinct account of the ' mathematical model,' he explains in some detail the attributes of relations and the postulates of order ; and this leads to an excellent discussion of the fundamental conditions of measurability and of the uses of measurement in psychophysics and other branches of experimental psychology. In an illuminating chapter on ' Selection,' Dr. Bechtoldt deals with problems involving correlational and weighting techniques (which are treated as including factor analysis) ; and the results of American factorial researches are referred to in other chapters, e.g., in the discussion of memory and motor skill. Nor are the contributions of British workers wholly neglected. Readers who are still sceptical of the validity of quantitative procedures in psychological enquiries may learn, from nearly every chapter, how the application of such methods, though never sufficient of itself, has led to new advances in the subject and to the more rigorous establishment of important generalizations in almost every field.

The main body of the work is divided into six sections, dealing in turn with physiological mechanisms, growth and development, motivation, learning and adjustment, sensory processes, and human performance. How far the arrangement and the contents mark an improvement on the older *Handbook* must be left to the experimentalist to decide. But the magnitude of the project is in itself a testimony to the enterprise and industry of the editor and his colleagues ; and there can be little doubt that, for some years to come, the work will remain the standard handbook in Britain as well as in America. *C. B.*

Essays in Psychology : Dedicated to David Katz. Uppsala : Almqvist & Wiksells. 1951. Pp. 284. 21*s.*

The essays contributed to this volume by eighteen eminent psychologists from nearly a dozen different countries show the high regard felt for Professor Katz and his work.

In comparing this collection with similar volumes published fifty years ago (for example, the *Psychological Essays* dedicated to William James), the statistical reader is immediately struck by the remarkable increase in the number of articles and investigations involving statistical techniques, for the most part dealing with problems of individual rather than of general psychology. This volume, after an article by Sir Frederic Bartlett on ' Anticipation in Human Performance,' starts with an historical account by Burt of the development of factor analysis and its main achievements ; Thurstone contributes a paper on ' Tests of Temperament ' ; Husen describes a factorial study of ' Personality as Revealed in Self Ratings ' ; there is a long paper by Ekman on ' The Number and Definition of Dimensions in Kretschmer's and Sheldon's Constitutional Systems,' and a short paper by Thouless on ' Individual Differences in Perception and their Significance in Psychology.' The names of Buytendijk, Carmichael, Meyerson, Michotte, Piéron, Révész, and Rosa Katz illustrate the high quality of the papers, and show that other fields have not been forgotten. Yet even here there is still a tendency to check the conclusions by a statistical test. The view recently expressed by certain British writers that " psychologists are coming more and more to agree that the devices of measurement, statistics, and mathematical techniques generally are not really applicable to the elusive processes of the mind " is evidently not shared by psychologists in other countries.

Readers of this *Journal* will find the volume full of suggestive papers, and will wish to join the contributors in sending their good wishes to Professor Katz. *E. R.*

Unravelling Juvenile Delinquency. By Sheldon Glueck and Eleanor Glueck. Pp. xvi + 400. London : Oxford University Press, 1950. £2 net.

The investigations carried out by Dr. and Mrs. Glueck on the efficacy of various methods of treating delinquents are well known to all students of criminology. Their findings have now led them to undertake a co-operative investigation into the problem of causation. The general plan is based on the double procedure first introduced by Burt when investigating causes of delinquency among London children, namely, making comparable case-studies of a fairly large sample of delinquents side by side with a carefully matched group of non-delinquents, and then applying a rigorous statistical analysis to determine which differences are significant.

In contrast to the majority of recent publications emanating from child clinics, the writers have paid special attention to the needs of an adequate statistical analysis. In this they have been assisted by Professor C. R. Doering, of Harvard University, who contributes a brief but suggestive discussion of the methods appropriate to this type of research. He concludes that for purposes like the present " over-refined statistical techniques are apt to become sterile " ; and hence he deems it prudent " to keep statistical computations to the minimum of complexity consistent with reliability and clarity." It is almost inevitable that both methods and results should be compared with Burt's. Unlike his enquiry, the present research makes no attempt to measure the relative strength or importance of the different causal factors by calculating coefficients of association or correlation. With quantitative data the investigators rely mainly on the calculation of means, standard deviations, and critical ratios. With qualitative data they have " found three class categorizations of qualitative material more reliable than either finer categories or a twofold classification " : thus their usual classification (like Burt's) distinguishes conditions as present in a marked or major degree, present in a minor degree, absent altogether. Significance is then tested by the chi-squared procedure.

It would be inappropriate here to discuss in detail the inferences drawn from this unusually thorough investigation. Briefly it may be said that, among their 500 delinquents, comparatively few could be regarded as showing any so-called ' psychopathic ' trend ; and, curiously enough, " neuroticism was more markedly present among the non-delinquents." Social and temperamental conditions were far more important than deviations in physical or mental health. In almost every case, however, the delinquency appeared to be the outcome, not of any single cause, but of a multiplicity of converging factors. It will be seen that the conclusions reveal a remarkable agreement with those already reached by British psychologists who have carried out similar investigations in London and elsewhere. *R. E. S.*

Measurement and Prediction. By S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, J. Clausen. New Jersey : Princeton University Press. Pp. xii + 756. 1950. $10.00.

This book is Volume IV of *Studies in Social Psychology in World War II.* The first two volumes, it may be recalled, dealt with ' The American Soldier,' and the third with ' Experiments on Mass Communication.' The last is essentially " a methodological volume," though (as the editor observes) the illustrative material has " an intrinsic substance interest."

The data on which all the studies have been based were collected by the Research Branch of the Information and Education Division of the U.S. War Department ; and the preparation of the series was rendered possible by a grant from the Carnegie Corporation of New York. The completion

of the present volume has been further assisted by financial grants from the Harvard University Laboratory of Social Relations and the Columbia University Bureau of Applied Social Research. In this country, as we have learnt from papers presented to the various psychological societies and sections, problems very similar to those discussed in the American series were also encountered by British psychologists who assisted the various Services ; and a vast mass of material must now be available in the numerous reports and memoranda which they drew up. Unfortunately, very little of this has so far been published, although it would be of the utmost value, not only to those responsible for the efficiency of the Services, but also to investigators interested in theoretical and practical questions of a more general type. British psychologists therefore will greatly envy the generous financial aid their American colleagues have received. The results, as they appear in this and other volumes, fully justify the expenditure ; and we may hope that their publication will do something to stimulate similar enterprise over here.

The general purpose of the book is to consider what statistical procedures are best suited for the analysis of qualitative data, particularly data obtained from questionnaires on attitudes, opinions, psychoneurotic symptoms, and the like. As Dr. Stouffer observes on the very first page, " the classification of the objects of study is vital to the development of social psychology as a scientific discipline " ; and the task of classification in its turn " requires rigorous yet economical methods for handling data which are initially *qualitative* and not quantitative." Among British writers, the exigencies of recent work, as well as the criticisms of the commoner quantitative procedures, have already led to a widening interest in the application of statistical methods to what Yule called ' classification by attributes ' and Pearson described as ' characters not quantitatively measurable.' Hence the detailed discussions in this volume will have an interest far beyond the more specialized types of problem with which it is primarily concerned.

The book falls into two main parts, dealing with questions of measurement and of prediction respectively. The first part consists of eleven chapters, largely theoretical in nature, and is concerned chiefly with the development of " models of ' ordered structures ' or scales," which shall be suited to material possessing a ' socio-psychological content.' After an introduction by the editor, giving ' an overview of contributions to scaling and scale theory,' Dr. Guttman takes up ' the problem of attitude and opinion measurement,' and then explains, in the course of seven illuminating chapters, the theory and application of the new technique which he has termed ' scale analysis.'

Dr. Guttman's aim is to find " an adequate basis for the quantification of qualitative data " ; and his first step is to determine whether any particular attitude area is really scalable. " The mass of current attitude research," as he justly complains, exhibits a " tendency to plunge into analysis of data without having any clear idea as to when a single dimension exists and when it does not." His own approach begins by " focusing on the ranking of individuals instead of on the ranking of items." He defines a set of items as forming a ' scale ' if all the persons who answer any given item favourably have higher ranks than those who answer it unfavourably. From this it follows that in a perfect scale " each person's responses should be reproducible from his rank alone." It is from this requirement that his " practical scalogram analysis " is derived.

Dr. Guttman, says the editor, turned in this " new and different direction because he saw clearly that the conventional methods of quantitative factor analysis, stemming from Spearman, were inappropriate for handling qualitative data." But a British psychologist may be tempted to ask whether earlier methods of factor analysis, stemming from Pearson and Yule, might not prove more appropriate. May we hope that some experienced factorist will attempt a systematic comparison of the two modes of approach ?

In practice, as Dr. Guttman points out, perfect scales are not to be expected : the best we can hope for is a ' quasi scale ' in which the errors are random. For such cases he proposes a ' coefficient of reproducibility ' to measure the degree to which the distribution obtained approximates to that of a ' perfect scale.' Having explained the theoretical basis, he then goes on to discuss in fuller detail his practical ' scalogram board technique,' the general utility of his method, the determination of reliability, and finally the ' principal components of scale analysis.' To readers of this *Journal* his last chapter will perhaps be of greatest interest. As they will have observed, earlier writers who regard factor analysis as primarily a method of classification (a view which Spearman expressly rejected) have repeatedly drawn attention to the systematic type of sign-pattern which is observable in nearly every factor-matrix derived from a well-balanced battery of test-items by simple summation. As a direct result of the mode of calculation, the pattern tends to exhibit what Burt calls a " hierarchical " scheme, and thus furnishes " a branching classification like that of the Porphyrian tree." [1] When obtained from a ' perfect scale,' the factors, as Ferguson has demonstrated, display these features in a particularly clear-cut fashion. " If," he says (referring to the passage just cited), " the (correlation) matrix is factorized by Thurstone's method or by the more efficient method described by Burt, all factors after the first are ' bipolar ' (i.e., they have negative as well as positive saturations), and are exemplary of the principle of dichotomous classification. The second factor-loading of any item is a measure of the deviation in difficulty of that item from the average difficulty of all

[1] *Factors of the Mind*, 1940, p. 309, and refs.

the items. The third factor represents a further dichotomization, each loading representing a deviation in difficulty from the average difficulty of those items that had been described as being either above or below the average difficulty by the second factor. The remaining factors represent further dichotomizations."

Guttman discovers much the same characteristics with his own method of analysis, and specifies them still more clearly. He shows that, when plotted, the graphs of the components so obtained can be expressed by polynomials of a degree corresponding with the number of their turning-points. What British psychologists would call the ' general ' factor he proposes to term the ' constant ' component, because in a perfect scale its saturations have the same value throughout. The first ' bipolar ' (Ferguson's ' second factor ') is, as he demonstrates, a " straight line function of the rank order " ; this he therefore terms the ' metric component.' The second bipolar (Ferguson's ' third factor ') is U-shaped ; and this, he suggests, may be called the ' intensity component.' The remaining components become increasingly oscillatory, each having one more turning-point than the last. It will be observed that, with Guttman's method of analysis, the rank-order of the individuals tested represents only one in the series of principal components, namely, the first bipolar. But he does not perhaps make it sufficiently clear that the ' components ' obtained by his procedure correspond to ' factors of difficulty ' (like the factors obtained by Ferguson), not to the more familiar ' factors of content.' [1]

Dr. Stouffer describes these chapters as representing a " drastic departure from earlier approaches." But he must be thinking solely of American approaches. In this country the problems raised by Dr. Guttman are by no means new. Thomson and his former students, notably Ferguson and Walker, have paid considerable attention to what they call ' answer patterns.' Thomson's definition of what he calls a ' unig test ' presents essentially the same concept as Guttman's definition of a ' pure scale,' and, as he has pointed out, implies " an *exact* relation between scores and answer pattern." What Thomson calls the ' coefficient of hig ' is an attempt to provide much the same kind of index as Guttman's ' coefficient of reducibility.' Walker has endeavoured to determine more precisely the conditions required if an answer pattern is to be perfectly reproducible ; and Ferguson has used matrix notation to deduce the peculiarities of such patterns and the product-sum matrices to which they give rise.[2]

Dr. Guttman devotes a chapter to the ' relation of scalogram analysis to other techniques,' such as ' non-scale approaches,' ' equal appearing intervals,' ' paired comparisons,' ' item-analysis,' and ' factor analysis.' Factor analysis, he states, " originated as a single factor theory by Spearman, and was developed into a multifactor theory by Thurstone and others " ; it is " designed only for quantitative variables." This is scarcely correct. Factor analysis originated with Pearson's proposal to transform a symmetrical table of product sums, correlations, or the like to terms of the ' principal axes ' of the multivariate frequency distribution. It was therefore from the start a ' multifactor theory.' Burt simplified the computations by introducing the ' product theorem ' and the ' simple summation ' formula, which were applicable to the contingency tables consisting of frequencies as well as to a table consisting of correlations or covariances. Spearman from the outset rejected the ' Pearson model ' as unsuited to work with mental tests, since these (in his view) required only a single common factor ; and eventually developed the ' tetrad difference method ' for demonstrating this conclusion. Thurstone's introduction of the ' centroid formula ' (identical with that for ' simple summation ') was in effect a return to the original procedure, with an attempt to supplement it by routine methods of computation and rotation.

As a basis for qualitative work Yule's formula for point correlations has achieved considerable popularity, though Yule himself preferred an alternative procedure. Guttman criticizes its use on the ground that when derived from a ' perfect scale ' the tetrads vanish only when all the elements are taken from the same side of the diagonal. This, however, had already been demonstrated by Ferguson [3] ; Yule's alternative formulæ do not seem open to this criticism. Perhaps the most convincing argument that Guttman brings forward turns on the point that " factor analysis is a vastly more tedious and expensive procedure." On this very practical ground he has certainly made out a strong case for his simpler methods. Moreover, from a theoretical standpoint, he has carried the discussion much further than his predecessors ; and the thorough and systematic way in which he

[1] See Gourlay, N., ' Difficulty Factors Arising from Tetrachoric Correlations,' this *Journal*, IV, 1951, pp. 65f., and Wherry, R. J., and Gaylord, R. H., ' Factor Pattern of Test-Items : Content, Difficulty, and Content Error Factors,' *Psychometrika*, IX, 1944, pp. 237f.

[2] The formula for the ' coefficient of hig ' is given by Walker in his paper on ' Answer Pattern and Score Scatter ' (*Brit. J. Psych.*, XXII, 1931, p. 78). Walker also suggests using ' Pearson's measure of goodness of fit.' Burt has suggested, as an alternative device, using the correlation ratio, which, when applied to ranks, yields a comparatively simple formula. Yet another possibility is to employ Kendall's ' coefficient of concordance ' (i.e., to find the ratio of the actual value of the sum of the squares of the deviations about the mean to the maximum value) ; but according to Burt this is identical with the correlation ratio.

[3] Ferguson, G. A., ' The Factorial Interpretation of Test Difficulty,' *Psychometrika*, VI, 1941, p. 326.

has deduced his conclusions from first principles is (as Dr. Stouffer says) both challenging and instructive.

Dr. Guttman's description of scale analysis is followed by Dr. Lazarsfeld's chapters on latent structure analysis. It will be seen that Dr. Guttman's deduction dispenses with the concept of an underlying continuum to which the responses are presumed to be related ; Dr. Lazarsfeld, on the other hand, " proposes a return to the older concept of a latent attitude continuum," and treats the probability that the respondents will mark the first (or positive) answer to any given item as a function of their position on this continuum. This function may be described by a curve which he calls the ' trace line.' His definition of a ' pure scale ' is expressed in terms of this function, and is obviously much wider than that of Dr. Guttman : a pure scale, he says, is formed by " an aggregate of items, such that the joint positive answers to any number of items have themselves a trace line which is the product of the original trace lines for the items viewed separately." Assuming that the ' trace lines ' can be expressed by polynomials, he shows that their degree will be one less than the rank of the matrix of joint frequencies for all possible pairs of items. " Guttman's quasi scale," he says, " can be derived analytically as a special case of latent structure analysis, Guttman's perfect scale becoming a limiting case of the quasi scale."

Dr. Lazarsfeld then takes particular cases for which certain simplifying conditions can be laid down, e.g., that the ' trace lines ' are linear, and that the ' population distribution ' is a point distribution. From these he derives working methods for determining the parameters of the distribution curve which will give ' optimum results.' If, for example, there are m items, and if the data are reducible to λ ' strata,' then the ' latent structure ' will be determined by $\lambda(m + 1)$ parameters. These he calls ' latent marginals for items.' He then discusses practical methods for computing these ' marginals,' and the application of the whole procedure to scale construction.

Once again it might be thought that much of the formulation could be described in the more familiar terms of factor analysis. Several of the basic theorems reached by Lazarsfeld seem identical with those given by Yule and Ferguson ; his ' latent attitude continua ' appear to be analogous to factors ; his ' latent marginals for items ' resemble the hypothetical saturations for tests that appear in the case of completely overlapping group factors ; and, since each item is stated to have its own specific component, the method involves the familiar device of reducing the ' communalities ' (i.e., the observed values in the leading diagonal). As we have seen, when a factor analysis is carried out with the usual summation procedure, the essential result is a set of bipolar factors, with negative as well as positive loadings. The obvious difficulty is then to transform these into factors with positive loadings only. Ferguson suggests substituting the triangular method of factor analysis which will represent the successive levels of difficulty measured by the tests (for an illustrative example, see this *Journal*, IV, p. 76, Table II B). But with qualitative data like that discussed by Yule (the presence or absence of certain physical characteristics), there can be no question of ' difficulty,' and in such cases Burt suggests rotating the bipolar-matrix, and using the frequencies of a higher order to derive the requisite rotation matrix direct from the data : this, as has been shown in several researches, can easily be done by employing Yule's equation for the ' case of complete independence ' (i.e., the case in which all the observed associations are attributable to one or more common factors). It would be illuminating to see these alternative procedures actually applied to the tables analysed by Lazarsfeld, and to determine how far the numerical solutions they provide resembled those which he obtains.

In his last two or three pages Lazarsfeld himself discusses the relation of latent structure analysis to factor analysis. Certain analogies, he says, might " make it appear that the present theory is nothing else than factor analysis where the form $[ij]$ plays the role of a correlation coefficient " (in the notation used by Ferguson $[ij] = p_{ij} - p_i p_j$). " But," he immediately adds, " this is not so." There is a clear difference between the two methods ; and, " as far as the writer can see, this is due to the fact that factor analysis assumes a *linear* relation between the factor and the test," whereas " the latent structure . . . may show a *curvilinear* relationship." In his approach, he says, " the notion of correlation is never needed : the theory can be based on the analysis of joint occurrence matrices." But once again there is no reason to assume that factor analysis is applicable only to correlations : the product moments can represent frequencies just as well as correlations or covariances.

In the concluding section we turn from measurement to prediction. Among these the two chapters of greatest interest to British psychologists will be those contributed by Dr. Shirley A. Star, which deal with the screening of psychoneurotics in the Army. The initial basis of the work consisted of responses to a " questionnaire covering 15 areas which the literature and psychiatric advice suggested would prove discriminating." Other types of test were also considered—e.g., the Rorschach ; but the results obtained appeared far less effective. The various stages in the construction of a workable questionnaire are described in some detail, and in the main follow the scaling procedures described in the preceding chapters. A striking discovery was that responses obtained from only *one* of the 15 areas originally included—namely, " a quasi scale of psychosomatic complaints "—proved " almost as discriminating as the entire battery of items covering all 15 areas." The paper-and-pencil test as finally developed became part of the regular routine at all induction stations in the United States during the last year of the war.

Dr. Star then analyses the figures secured (*a*) from the screening tests and (*b*) from the psychiatric examinations with a batch of over 100,000 men. During the development of the test, the two sets of data were kept quite independent ; with this larger group the test scores were available for the examining physicians to refer to if they wished. The amount of efficiency possessed by the final version is roughly indicated by the fact that, with a fairly high borderline (cutting off one-third of the recruits examined), those screened out by the test included nearly 70 per cent. of those rejected later by a psychiatric examination. This result is even more promising than the figure obtained might at first sight suggest, because the ultimate criterion—namely, rejection after an individual examination by a psychiatrist—proved far from trustworthy. Commenting on the data supplied by the psychiatrists Dr. Stouffer rightly observes that the variations from one station to another " will seem almost fantastic." The proportions of men rejected ranged from 0·5 per cent. at one station to 51 per cent. at another ; the proportion of the rejects classified as ' psychoneurotic ' ranged from 2·7 to 90·2 per cent. ; and the proportion of those classified as ' psychopathic ' from 0 to 81·3 per cent. Judged by the tests, the actual differences between the groups examined at different stations were comparatively slight. Thus, as the authors conclude, " when the psychiatric diagnoses were so different one cannot expect to find the test very efficient in predicting psychiatrists' behaviour ; until psychiatric diagnosis is better standardized than it is to-day, it is likely that the predictability of such screening tests will appear to fall far short of the ideal." *A. R. S.*

The Scientific Study of Personality. By H. J. Eysenck. London : Routledge and Kegan Paul, 1952. Pp. xiv + 320. 30*s*.

In this book, the author tells us, he is " concerned with the problem of taxonomy in psychology," that is, with " the question of *classifying* traits and types " and presumably persons—particularly neurotic or psychotic persons. His primary aim is " to discover the main dimensions of personality, and to define them by means of strictly experimental quantitative procedures." This, indeed, was the purpose of his previous book on *Dimensions of Personality*. But in the present volume " an attempt is made to advance more rigorous proofs."

Even those readers who may not agree with Dr. Eysenck's conclusions will be compelled to admire the ingenuity and industry which his investigations display, and the lively and challenging fashion in which his conclusions are set forth. However, so many of his contentions involve what Professor Lewis describes in his Foreword as " a root and branch attack " on the methods of other workers in the same field, that both psychologists and psychiatrists will be tempted to wonder how far the somewhat technical procedures adopted—especially the statistical analyses—possess the merits alleged. This part of the work therefore seems to call for a somewhat detailed examination.

As its title indicates, the book seeks above all to present a ' scientific ' study. Others have made a similar claim ; but, says Dr. Eysenck, " finding their contributions unacceptable as ' scientific ' in the orthodox sense, they have surreptitiously altered the meaning of the word, so as to retain its prestige value." Accordingly, in the opening chapter he begins by explaining what in his view should be the basic nature of any genuinely scientific attempt. Citing the work of Newton and Einstein as offering ideal models, he endeavours to show that all scientists adopt the " hypothetic-deductive method." This involves four steps. Thus, in the psychological field (to take his own example), the scientific investigator must first formulate " a clear unambiguous hypothesis " (e.g., that one dimension of personality is neuroticism). Secondly, " deductions, preferably quantitative, are made " : for instance, if the hypothesis just specified is true, it should follow that there is a factor common to all appropriate tests, whose saturations correlate highly with their correlations with an independent criterion. Thirdly, taking as an independent criterion the classification of the persons tested into normal and neurotic, experiments are carried out in the hope of verifying this corollary. When that is done, we find (he says) a distinctly high correlation, namely, one in the neighbourhood of 0·80. Finally, therefore, we conclude that our hypothesis is confirmed.

This is in effect a reversion, perhaps unconscious, to J. S. Mill's account of the ' deductive method ' in physical science. Mill himself, however, held that it was not appropriate to the study of human nature. Now, as has been pointed out by nearly every logician since, this mode of reasoning manifestly involves a formal fallacy : " If neuroticism exists, such-and-such a correlation would be found. We do find it. Therefore neuroticism exists." Such an inference is exactly parallel to the famous argument attributed to Judge Jeffreys : " If Mistress Gaunt had been guilty, you may rest assured she would have been burned. She was burned. *Ergo*, she was guilty." [1]

[1] Burt (*Factors of the Mind*, 1940, pp. 33f.) puts the matter, as he says, " succinctly if somewhat pedantically " : the alleged inference follows " the *modus ponendo ponens*, and commits the fallacy of ' affirming the consequent,' whereas it ought to follow the *modus tollendo ponens*, and proceed by disproving all the alternatives except the one desired." Criticizing this mode of reasoning in factorial psychology, he remarks : " Hume, Adam Smith, Bentham, and James Mill all explicitly took Newton's ' hypothetico-deductive method ' for their model, and so reached conclusions about human motives, which were (as even Macaulay was quick to point out) far removed from *actual* motivation. Better models would be found in chemical and biological researches, e.g., those of Boyle, Harvey, Lavoisier, or Pasteur. Even in physics, as Boyle long ago observed in his comments on Pascal's ' pneumatics,' planned experiments are needed to *disprove* the traditional hypothesis as well as to confirm the new one."

The point is still so continually overlooked by writers on factor analysis that it calls for some examination here, particularly as Dr. Eysenck lays repeated stress upon this ' method ' throughout his volume. The classical instance of the fallacy is the original argument for the ' two-factor theory.' " If cognitive tests involve only a single general factor and a set of specific factors peculiar to each, the table of correlations will be hierarchical. Within the margin imposed by sampling errors, this table, obtained by experiment, is hierarchical. Therefore the results of such tests are due simply to one common factor and a specific." But, as Thomson quickly pointed out, at least one alternative hypothesis is equally tenable, namely, that the results of the tests and their correlations are the effect of a number of overlapping group factors. Dr. Eysenck seeks to support his account by a long quotation (or rather a series of extracts run together) from Professor Stebbing's *Introduction to Modern Logic*. But one of the sentences he puts within inverted commas—" the method of science . . . has four essential stages "—does not occur in her discussion ; the ' stages ' quoted are not given by her as a description of the ' method of science,' but of ' commonsense thinking ' on which science improves. Moreover, Dr. Eysenck stops short before the important clause with which Professor Stebbing ends : " It would be felt that, in the given case, no *other* supposition would so well account for all the observed facts " (her italics). Later she quotes Jevons as saying : " a crucial experiment must not simply confirm one theory, but must *negative another ;* it must decide between two equally plausible views."

Thus, so far as it is possible to generalize about ' scientific ' inferences, the form of reasoning underlying the most convincing enquiries would seem to be, not ' hypothetico-deductive,' but ' disjunctive-deductive.' [1] This is true whether our statistical arguments follow the Pearsonian scheme derived from Bayes' theorem or whether they follow Fisher's *Design of Experiments ;* (the need for an initial disjunctive premiss is clearly brought out in the model experiment in sensory discrimination with which Fisher starts). Now it may be reasonably claimed that throughout his book many of Dr. Eysenck's inferences fail to carry conviction just because he has overlooked the need for the disjunctive premiss. To prove his hypotheses it is not sufficient to report experiments that are consistent with his corollaries : *he should also have attempted to disprove the alternatives.* When he does notice alternatives, they are usually dealt with by the short and easy method of " abusing the defendant's attorney." Thus psychologists and psychiatrists who support earlier schemes of classification are accused of donning the worn-out ' Kraepelinian straightjacket ' [*sic*], or, when they have thrown this off, of ' indulging in Freudian manic spells.'

The ' hypothetico-deductive ' principle leads, we are told, to a new factorial method which Dr. Eysenck calls ' criterion analysis.' This novel device is needed because " the resolution of a given matrix of inter-correlations can be carried out in an infinite number of ways, a difficulty which is not faced at all by some writers, like Burt or Stephenson." [2] And the greater part of his book, he tells us, is " devoted to an application of this method." But is this further proposal anything more than a reaffirmation of the principle so often urged by Burt himself, namely, that " where possible, the interpretation of the internal criterion or ' factor ' should be checked by comparison with an independent or external criterion ; . . . when required, a numerical comparison of the two can always be obtained by correlating the measurements for the factor with the measurements for the criterion " ? There is perhaps a minor difference. Dr. Eysenck prefers to correlate, not the measurements, but the test-saturations for the factor with the test-correlations for the criterion ; but he offers no algebraic proof for this somewhat questionable change. And in order to secure a " position for the factor which is both unique and invariant," he carries out a rough graphical rotation to bring the factor into " that position for which the correlation between the factor and the criterion is a maximum." One would have supposed that in doing this the investigator would eventually approach a position which was identical with that of the criterion-axis, and that the maximized correlation (if directly calculated instead of being derived from the graph) would be simply the ordinary multiple correlation between the criterion and the tests. It is true that he goes on to say that we may also increase the correlation by improving the criterion, much as has been done by improving " teachers' ratings of intelligence." But at this point the procedure would cease to be a statistical technique, and become a method of improving observational techniques for assessing what is primarily an observational concept. How this is to be done we are never told.[3]

[1] It is hardly necessary to cite additional authorities : perhaps the clearest statement is that of Joseph (*Introduction to Logic*, pp. 430f.) who gives numerous illustrations from the biological sciences.

[2] This last clause is a little puzzling because Burt has not only " faced " the difficulty, but suggested solutions : e.g., if we are seeking bipolar factors, a unique resolution is obtained by seeking factors in order of their contribution to the total variance : (' simple summation ' is used in practice as yielding a convenient and sufficiently close approximation to such factors) ; if we are seeking group factors, a resolution based on the lines of demarcation indicated by the bipolar factors is both unique and (when other tests are added) invariant.

[3] Dr. Eysenck's proposal has already been critically examined in a paper by Dr. Ardie Lubin : but no mention is made of these criticisms, nor is any defence attempted. No doubt other readers will offer their views on this technique, to which the author attaches great importance in almost every chapter.

The fact is that the hypothesis of an innate general ability, approximately identifiable with what the teacher calls ' intelligence,' owes its plausibility, not to *one* conclusive argument, whether deductive or inductive, nor yet to the comparison of two types of evidence (e.g., identifying the internal criterion derived by statistical analysis with an independent concept derived from the cumulative observations of experienced teachers). " Like most well-established theories in science, it is a resultant of a large number of converging lines of evidence : it forms the core of a systematic conceptual scheme, which claims acceptance, partly because it serves to co-ordinate and organize a wide variety of facts and generalizations (often reached by many independent investigators), and partly because it serves, and has served, as a most fruitful source of new corollaries and new problems for research." [1]

Thus, in his anxiety to commend what he calls " *our* hypotheses," and to decry those of all other writers as scarcely worthy of notice, Dr. Eysenck only succeeds in cutting from beneath his feet some of the most valuable grounds of support : namely, the cumulative and conjoint evidence furnished by independent investigators who have already examined very similar problems and come to very similar conclusions.

Let us now turn from the discussion of general methodology to the nature of the particular hypotheses themselves. The most fundamental is the assumption that the structure of personality is essentially " atomistic and elementalistic." This, as he admits, is a hypothesis which some of the most eminent writers on personality, like Allport and Windelband, have vigorously contested : they favour the opposite assumption, the ' wholistic,' ' Gestaltist,' ' organismic ' hypothesis, as it is variously termed. Every personality, they maintain, is unique, and therefore " cannot possibly be analysed into an aggregate of numerical measurements." However, this alternative assumption is disposed of in his usual blunt fashion. " It is undeniable," says Dr. Eysenck, " that Professor Windelband is unique : so is my old shoe." Unique or not, individuals certainly differ ; and they " cannot be meaningfully said to differ unless they are compared along some quantitative variable."

But this contention implies a further assumption that seems equally open to question. Many psychiatrists would reply that " the difference between a man who is, and a man who is not, suffering from general paralysis is a difference, not of degree, but of kind, and is therefore to be expressed in terms, not of a quantitative variable, but of qualitative attributes. It is therefore conceivable that, if factorial methods are to be applied to pathological cases, what is needed is not an analysis of the more familiar type, based on quantitative measurements, but an analysis applicable to qualitative differences, like that used by Burt for distinguishing neurotics and delinquents, or clinical conditions turning on the presence or absence of certain physical symptoms. Dr. Eysenck, however, prefers to assume that " there exists a ' neuroticism ' continuum and a ' psychoticism ' continuum " ; then, applying tests that yield measurements on a continuous scale, and subjecting the measurements to a factor analysis, he reaches results that are in harmony with a hypothesis of continuity, and therefore claims to have " shown " that such continua exist : thus we are told that patients belonging to the schizophrenic group are " somewhere between the normal and the manic depressive group," and that all such groups can be assessed in terms of the same psychotic dimension.

Having decided to assume continuity, his next step is to consider what kind of observational data are to be procured for the purpose of the projected factorial investigations. Dr. Eysenck argues that " the only type of data lending themselves to analysis of the kind proposed are those collected through the use of tests " : observational methods or clinical procedures he considers scarcely worth discussion. " Two techniques," he says, " are used almost exclusively in current practice—the questionnaire and the so-called projective tests " (a rather surprising statement). He himself decides " to reject the so-called projective techniques, and concentrate on tests which have been much neglected lately, viz., ' objective behaviour tests,' " that is, tests based on what other observers have called the ' method of expression ' (e.g., dexterity, persistence, tapping, body-sway, etc.).

The usual objection to such tests is that their reliability is so low, and that their validity cannot possibly exceed their reliability. Unfortunately Dr. Eysenck's tables seldom include reliability coefficients ; and the few that he gives are by no means high. He is apparently unaware of the fairly comprehensive survey of the reliability and validity of different methods brought together in the Symposium on Personality and published a few years ago (*Brit. J. Educ. Psychol.*, ' The Assessment of Personality,' XV, 1945, pp. 115f.). The conclusions then reached were that laboratory tests depending on the ' method of expression ' are less reliable and less efficient than those depending on the ' method of impression ' (camouflaged questionnaires, apperception tests, association tests, picture tests, and the like) ; that tests are in general less trustworthy than assessments reached by trained and experienced observers using methods of interviewing, particularly when the persons to be studied are mentally abnormal ; that assessments based on observed behaviour in standardized real-life situations are better still, and the pooled results of a combination of procedures best of all. Though Dr. Eysenck would not accept them, there is little in his figures to contradict these conclusions. For example, in the biggest of his tables, out of 378 correlations only 5 are numerically larger than

[1] Burt, *loc. cit.* This is the feature by which Professor Stebbing distinguishes ' the method of science ' from ' ordinary commonsense thinking ' : the former, she says, " achieves a far higher degree of organization "—a point which Dr. Eysenck's quotations fail to bring out.

± ·500, and most of these are for tests of similar content. In the chapter on ' objective measurement ' he refers the reader for fuller details to the theses of Dr. A. Clarke and Dr. A. Gravely, and we are told that their account " has been followed very closely." Yet, on looking up their own statements, we find that Dr. Clarke concludes that " none of the tests has a sufficiently high validity coefficient to claim any practical value," and Dr. Gravely adds that there is little evidence that the non-verbal (i.e., ' objective ') tests yield better indications than others.

Dr. Eysenck is severely critical of " the mole-like calculators of innumerable correlations, who seem incapable of seeing the wood for the trees." Yet a good deal may be learnt about the wood by carefully observing the trees. In the study of neurotic and delinquent children, some of the most fruitful discoveries have been made by those who combined observational methods with tests and statistics. With so many trained psychological workers to assist in the investigations, it is curious that no attempt was made to secure introspective evidence about the nature of the mental processes which the tests actually involve. And, at a mental hospital, where patients are available for prolonged observation by experienced medical psychologists, the author might conceivably have hit on more suggestive hypotheses and better diagnostic methods had he made use of their first-hand observations. As one psychiatric critic has already observed, he seems to have derived his initial ideas more from a study of periodicals than from a study of patients.

Finally, let us consider the main conclusions derived from the statistical analyses. In describing ' the dimensional approach ' he begins with ' the general picture of the organization of mental (or cognitive) abilities.' Here, as a result of factor analysis, it is claimed that we can " discern four main levels." These are set out in a suggestive hierarchical pedigree table with general intelligence (or *g*) at the top, and various ' primary mental abilities ' (cognitive group factors) branching out. But the most important part of his book is devoted to establishing three non-cognitive factors of ' personality ' : namely, (i) ' a general factor of neuroticism,' (ii) a bipolar factor representing extraversion versus introversion, and (iii) a third factor called ' psychoticism.' All these, it is said, are ' approximately orthogonal ' to the cognitive factor of intelligence, ' operationally defined in terms of Thurstone's second-order factor.'

The factorial scheme thus reached inevitably suggests a comparison with those published by other investigators. Indeed, Dr. Eysenck's previous volume stressed his indebtedness to " Burt's writings," and pointed out that his results led to " a hierarchical view of personality structure, akin to Burt's." Now, however, he would seem to be of a very different view. " Burt's system of classification " (he says) " is some thirty years out of date, and discredited in contemporary psychology for many valid reasons. Burt " (he adds) " has recently taken up this notion again, and has contrasted the factor of emotionality with the factor of neuroticism." In reply Dr. Eysenck now maintains " there is no support at all in favour of the view that general emotionality has any claim to be regarded as a strongly marked and important trait of personality." His objection seems to be that the earlier work, leading to the analogous factors, was based solely on impressionistic ratings obtained from teachers. But in point of fact the very attempt at factorizing non-cognitive traits began by criticizing such ratings as are ordinarily obtained ; instead, Burt tells us, the correlations were based on " items of behaviour systematically recorded " (*Brit. Ass. Ann. Rep.*, 1915, p. 695) ; and in his recent discussion he does *not* " contrast the factor of general emotionality with the factor of neuroticism " : he suggests that the two are virtually *identical*, and merely deprecates the new label on the ground that not every highly unstable person necessarily succumbs to a neurotic disorder. Later on in his book Dr. Eysenck himself refers to this factor as a factor for ' emotional instability ' ; and thus seems unconsciously to recognize the identity of the two.

There are, indeed, several other features which increase the similarity between the two systems. He states that in his earlier studies " factor analysis was used to establish two main factors in body build : (i) a general factor of body size and (ii) a type factor " (roughly distinguishing the lean or ' long ' from the ' broad ') ; he finds that " the neurotics are more prone to the lean type," and that the " dysthymics (extreme introverts) were also markedly leptomorph." Now, as it happens, in the applications of factor analysis to the study of body types, carried out over many years by Burt and his students (Cohen, Hammond, Banks, and others), precisely the same two factors appeared ; and in Burt's 1937 paper (which Dr. Eysenck dismisses as " discredited ") it was noted that the correlations obtained for both children and adults revealed a " tendency for the neurotic or introverted to be thin or lean." The only other ' general factor ' to which Dr. Eysenck refers is the ' general factor of æsthetic judgement.' Its ' demonstration ' is again said to be due to Dr. Eysenck. Yet it figures prominently in the various versions of the scheme he quotes as ' out of date.'

However, one of the more specific ' dimensions ' which Dr. Eysenck claims to have established is certainly new, at least as a statistical ' factor,' namely, the dimension he terms ' psychoticism.' The chapter dealing with this and with the interesting points raised by Kretschmer is the most valuable portion of the book. On various grounds the hypothesis of a general ' psychopathic diathesis ' has been championed by several early psychiatrists, and strongly criticized by others. But once again these earlier discussions are somewhat contemptuously dismissed. Whether or not psychiatrists will be convinced by the evidence here offered, it cannot be denied that both the data obtained and the detailed discussion are of outstanding interest. Here, indeed, is a contribution that is both original

and suggestive, and well deserves to be followed up. On the other hand, it is surprising to find no mention whatever of the factors reported by an energetic group of workers inspired by Spearman and Bernard Hart. After all, their work at the Maudsley, Bethlem, and Horton Mental Hospitals represents the earliest and (in the view of many) one of the most suggestive attempts to apply factorial techniques to the study of patients in such institutions.[1]

In his closing chapter on the ' Organization of Personality ' Dr. Eysenck gives his final grounds for rejecting the ' organismic ' view. He quotes the work of Kelley and Fiske as demonstrating that " adding scores is superior to the organismic view of organizing data into a meaningful pattern " ; and describes a recent investigation carried out at his suggestion by Granger on colour preferences where preferences predicted on the ' elementalistic ' hypothesis were confirmed. He concludes by saying " it is difficult to see how the Gestalt hypothesis could be invalidated more conclusively." Yet once again, although he has no difficulty in showing that the results are fully consistent with his own hypothesis, this of itself does not disprove the alternative hypothesis. As a matter of fact, there is no reason why both should not be true. Finally, a ' bibliography ' of over 300 numbers is appended, which future investigators may usefully consult, if only because of the wide acquaintance it shows with comparatively unfamiliar publications in Germany and America. British articles on the subject (except by himself or his colleagues) are mainly ignored : readers of this *Journal*, for example, will recall at least a dozen articles bearing closely on his problems which do not figure in his list.

In a statistical journal it seems proper to concentrate mainly on methods and results of interest to the statistical psychologist. But it would be grossly unfair to make no reference, if only quite briefly, to the other topics treated and to the undoubted merits of the work. Having established his factors, Dr. Eysenck goes on to discuss various practical ' applications of dimensional research,' which are full of interest to psychological workers as well as to psychiatrists. And throughout the volume the ingenuity, industry, and enthusiasm displayed should, let us hope, stimulate those who disagree with him to produce factual rather than speculative evidence, and so to contribute to the advancement of what perhaps may be christened factorial psychiatry. *W. L. G.*

[1] See, for example, the seven papers on ' Spearman Factors and Psychiatry ' published by Dr. Stephenson, Dr. Murdo Mackenzie, Miss Simmins, and others (*Brit. J. Med. Psychol.*, XIV, 1934, pp. 101–35).

INDEX TO VOLUME V

INDEX OF AUTHORS

Authors	*Titles*	*Page*
Banks, C., and Keir, G.	A Factorial Analysis of Items in the Bernreuter Personality Inventory	19–30
Burt, C., and Howard, M.	The Nature and Causes of Maladjustment among Children of School Age	39–58
Burt, C.	Tests of Significance in Factor Analysis	109–133
Guilford, J. P., Michael, W. B., and Perry, N. C.	The Estimation of a Point Biserial Coefficient of Correlation from a Phi Coefficient	139–150
Hamilton, M.	An Iterative Method for Computing Inverse Matrices	181–188
Hick, W. E.	The Entropy of an Ensemble which is Divided into Classes	197
Howard, M., and Burt, C.	The Nature and Causes of Maladjustment among Children of School Age	39–58
Keir, G., and Banks, C.	A Factorial Analysis of Items in the Bernreuter Personality Inventory	19–30
Kestelman, H.	The Fundamental Equation of Factor Analysis	1–6
Mellone, M. A., and Roberts, J. A. F.	On the Adjustment of Terman-Merrill I.Q.s to Secure Comparability at Different Ages	65–79
Michael, W. B., Perry, N. C., and Guilford, J. P.	The Estimation of a Point Biserial Coefficient of Correlation from a Phi Coefficient	139–150
Moursy, E. M.	The Hierarchical Organization of Cognitive Levels	151–180
Neuhaus, J. O., and Wrigley, C.	A Re-factorization of the Burt-Pearson Matrix with the Ordvac Electronic Computer	105–108
Perry, N. C., Michael, W. B., and Guilford, J. P.	The Estimation of a Point Biserial Coefficient of Correlation from a Phi Coefficient	139–150
Pilliner, A. E. G	The Application of Analysis of Variance to Problems of Correlation	31–38
Renshaw, T.	Factor Rotation by the Method of Extended Vectors	7–18
Roberts, J. A. F., and Mellone, M. A.	On the Adjustment of Terman-Merrill I.Q.s to Secure Comparability at Different Ages	65–79
Sanai, M.	An Empirical Study of Political, Religious, and Social Attitudes	81–92
Sutherland, J.	The Prognostic Value of Certain Arithmetic Tests at the Age of Eleven Plus	189–196
Wrigley, C.	The Prediction of a Complex Aptitude	93–104
Wrigley, C., and Neuhaus, J. O.	A Re-factorization of the Burt-Pearson Matrix with the Ordvac Electronic Computer	105–108
Correspondence	The Estimation of Communalities	59
	Tetrachoric Correlations and Difficulty Factors	60
	Bipolar Factors as a Cause of Cyclic Overlap	197–202
Obituary	Dr. William Brown	137–138

INDEX OF SUBJECTS

Titles	*Authors*	*Page*
Analysis of Variance, The Application of, to Problems of Correlation	Pilliner, A. E. G.	31–38
Arithmetic Tests at the Age of Eleven Plus, The Prognostic Value of Certain	Sutherland, J.	189–196
Attitudes, Political, Religious, and Social, An Empirical Study of	Sanai, M.	81–92
Bernreuter Personality Inventory, A Factorial Analysis of Items in the	Banks, C., and Keir, G.	19–30
Bipolar Factors as a Cause of Cyclic Overlap (Correspondence)	Leclère, H.	197–202
Brown, Dr. William (Obituary)		137–138
Burt-Pearson Matrix, A Re-factorization of the, with the Ordvac Electronic Computer	Neuhaus, J. O., and Wrigley, C.	105–108

INDEX OF SUBJECTS—*Continued*

Titles	*Authors*	*Page*
Communalities, The Estimation of (Correspondence)	Dixon, A. L.	59
Complex Aptitude, The Prediction of a	Wrigley, C.	93–104
Difficulty Factors, Tetrachoric Correlations and (Correspondence)	Warner, E. N.	60
Entropy of an Ensemble which is Divided into Classes, The	Hick, W. E.	197
Factor Analysis, Tests of Significance in	Burt, C.	109–133
Factor Analysis, The Fundamental Equation of	Kestelman, H.	1–6
Factor Rotation by the Method of Extended Vectors	Renshaw, T.	7–18
Hierarchical Organization of Cognitive Levels, The	Moursy, E. M.	151–180
Inverse Matrices, An Iterative Method for Computing	Hamilton, M.	181–188
Maladjustment among Children of School Age, The Nature and Causes of	Burt, C., and Howard, M.	39–58
Point Biserial Coefficient of Correlation, The Estimation of a, from a Phi Coefficient	Michael, W. B., Perry, N. C., and Guilford, J. P.	139–150
Terman-Merrill I.Q.s, On the Adjustment of, to Secure Comparability at Different Ages	Roberts, J. A. F., and Mellone, M. A.	65–79

BOOKS REVIEWED

Authors	*Titles*	*Page*
Barlow, F.	Mental Prodigies	64
Eysenck, H. J.	The Scientific Study of Personality	208–212
Glueck, S., and Glueck, E.	Unravelling Juvenile Delinquency	204
Guilford, J. P.	Fundamental Statistics in Psychology and Education	134
Gulliksen, H.	Theory of Mental Tests	134–135
Kirman, B. H.	This Matter of Mind	135–136
Lindquist, E. F.	Educational Measurement	203
Piéron, H.	Vocabulaire de la Psychologie	64
Stevens, S. S.	Handbook of Experimental Psychology	203
Stouffer, S. A., *et al.*	Measurement and Prediction	204–208
Zangwill, O. L.	An Introduction to Modern Psychology	61-63
—	Essays in Psychology : Dedicated to David Katz.	204

TP
UNIVERSITY OF LONDON LIBRARY
DO NOT REMOVE

Vol. V Part I March 1952

THE BRITISH JOURNAL OF PSYCHOLOGY

Statistical Section

EDITED BY
CYRIL BURT AND GODFREY THOMSON

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ARTHUR SUMMERFIELD
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
B. BABINGTON SMITH
W. STEPHENSON
P. E. VERNON
J. W. WHITFIELD

Managing Editor: H. G. MAULE

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4.

Price Fifteen Shillings

CONTENTS

		Page
KESTELMAN, H.	The Fundamental Equation of Factor Analysis . .	1
RENSHAW, T.	Factor Rotation by the Method of Extended Vectors .	7
BANKS, C., and KEIR, G.	A Factorial Analysis of Items in the Bernreuter Personality Inventory	19
PILLINER, A. E. G.	The Application of Analysis of Variance to Problems of Correlation	31
BURT, C., and HOWARD, M.	The Nature and Causes of Maladjustment Among Children of School Age	39
CORRESPONDENCE		
(1)	The Estimation of Communalities	59
(2)	Tetrachoric Correlations and Difficulty Factors . .	60
BOOK REVIEWS	. . . . .	61

The British Journal of Psychology (*Statistical Section*) is issued by the British Psychological Society. Three numbers will be published annually.

The subscription price is 30*s.* a volume of about 200 pages. Subscriptions should be sent to any Bookseller, or to THE MANAGER, University of London Press, Ltd., Warwick Square, London, E.C.4.

Papers for publication should be sent either to Sir CYRIL BURT, 9 Elsworthy Road, London, N.W.3, or to Sir GODFREY THOMSON, 5 Ravelston Dykes, Edinburgh, 4. Contributors will receive 25 copies of their papers free : extra copies may be had at cost price, if the order is given when the proofs are returned.

Authors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the manuscript (including a brief list of contents at the head and a summary of results and conclusions at the end) should follow the general arrangement observed in the publications of the British Psychological Society.

Books for review should be sent to one of the Editors ; advertisements for insertion to the Manager, University of London Press, Ltd.

The British Psychological Society also issues quarterly *The British Journal of Psychology, General Section*, and *The British Journal of Medical Psychology*. Subscriptions, 40*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society also publishes, jointly with the Training College Association, *The British Journal of Educational Psychology*. Subscriptions, £1 per annum, for this Journal should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2.

Papers for publication in the *General Section* should be sent to Professor D. W. HARDING, the Psychological Laboratory, Bedford College, London, N.W.1 ; those for publication in *The British Journal of Medical Psychology* should be sent to Dr. J. D. SUTHERLAND, 2 Beaumont Street, London, W.1 ; those for publication in *The British Journal of Educational Psychology* to Emeritus Professor C. W. VALENTINE, The White House, Wythall, Nr. Birmingham.

1956 II
1958 I